REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-01-

0370

1. REPORT DATE *(DD-MM-YYYY)* 29-03-01	**2. REPORT TYPE** Final Technical Report	**3. DATES COVERED** *(From - To)* 2/15/98 - 12/31/00

4. TITLE AND SUBTITLE SHARED-KNOWLEDGE AND TEAM PERFORMANCE: A COGNITIVE ENGINEERING APPROACH TO MEASUREMENT	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** F49620-98-1-0287 **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Nancy J. Cooke, Steven M. Shope, Preston A. Kiekel	**5d. PROJECT NUMBER** 2313 **5e. TASK NUMBER** BV **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) New Mexico State University Office of Sponsored Programs, MSC 3CRL PO Box 30001 Las Cruces, NM 88003-8001	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) John Tangney AFOSR/NL 801 North Randolph Street, Room 732 Arlington VA 22203-1977	**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED
A

20010625 168

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A three-year research effort is described in which a synthetic team task was developed in the context of Uninhabited Air Vehicle operations. The synthetic task was abstracted from actual team operations of Air Force's Predator, guided by multiple research and pragmatic constraints. This synthetic environment, including a number of custom-designed experimental control and data collection measures and tools, provided a backdrop for various methodological developments and research on team cognition and its relation to team performance.

Team cognition can be viewed as the collective cognition of the individual team members as processed by team behaviors such as communication and coordination. A number of measurement issues were identified relevant to this perspective. In this light, measures and metrics of team knowledge and team situation awareness were developed and evaluated in two empirical studies in the synthetic environment. Results indicate that a measure of team taskwork knowledge based on relatedness ratings and a query-based measure of team situation awareness were predictive of team performance differences. Further, patterns of team skill acquisition and effects of a training intervention were examined. Findings support the premise that the synthetic environment provides a rich and complex test bed for future research on team cognition.

15. SUBJECT TERMS

Team performance, team cognition, measurement, synthetic task

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Nancy J. Cooke
a. REPORT Unclas	**b. ABSTRACT** Unclas	**c. THIS PAGE** Unclas	UL	68 + 25 appendix 93	**19b. TELEPHONE NUMBER** *(include area code)*505-646-1630

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

FINAL TECHNICAL REPORT

SHARED-KNOWLEDGE AND TEAM PERFORMANCE: A COGNITIVE ENGINEERING APPROACH TO MEASUREMENT

Nancy J. Cooke, Steven M. Shope, Preston A. Kiekel

New Mexico State University

Las Cruces, NM

Contact Information:

Nancy J. Cooke
Department of Psychology
New Mexico State University
Las Cruces, NM 88003

Voice: 505-646-1630
Fax: 505-646-6212
Email: cooke@crl.nmsu.edu

ABSTRACT

A three-year research effort is described in which a synthetic team task was developed in the context of Uninhabited Air Vehicle operations. The synthetic task was abstracted from actual team operations of Air Force's Predator, guided by multiple research and pragmatic constraints. This synthetic environment, including a number of custom-designed experimental control and data collection measures and tools, provided a backdrop for various methodological developments and research on team cognition and its relation to team performance.

Team cognition can be viewed as the collective cognition of the individual team members as processed by team behaviors such as communication and coordination. A number of measurement issues were identified relevant to this perspective. In this light, measures and metrics of team knowledge and team situation awareness were developed and evaluated in two empirical studies in the synthetic environment. Results indicate that a measure of team taskwork knowledge based on relatedness ratings and a query-based measure of team situation awareness were predictive of team performance differences. Further, patterns of team skill acquisition and effects of a training intervention were examined. Findings support the premise that the synthetic environment provides a rich and complex test bed for future research on team cognition.

TABLE OF CONTENTS

RESEARCH TEAM

The research and development work reported in this document was carried out and influenced by the following extensive team of faculty, students, contractors, consultants, and colleagues.

Principal Investigator, New Mexico State University: Nancy Cooke
AFOSR Program Managers: Willard Larkin, John Tangney
CERTT Lab Developer, Sandia Research Corporation: Steven Shope, Richard Crawford, Evelyn Crawford, Gary Noble
USAF Colleagues Brooks AFB and Mesa, AZ: Linda Elliott, Leo Gugerty, Ellen Hall, Elizabeth Martin, Sam Schifflett
Graduate Assistants, New Mexico State University: Greg Bromgard, Amy Burgess, Sara Caukwell, Anne Duran, Erin Helm, Steve Jackson, Heather Jiron, Preston Kiekel, Rebecca Keith, Honore Lassiter, Skye Pazuchanics, Harry Pedersen, Krisela Rivera
Undergraduates, New Mexico State University: Gary Bond, Celine Samaniego, Natasha Sutherlin, Bien van der Meulen
Consultant, University of Central Florida: Clint Bowers
Air Force ROTC Coordinator: Terry Adler
Army ROTC Coordinator: William Dorman
Colleague, New Mexico State University, Management Department: Janice Black

INTRODUCTION

THE PROBLEM

Technological developments in the military and elsewhere have transformed highly repetitive manual tasks, requiring practiced motor skills to tasks that require cognitive skills often related to overseeing new technology such as monitoring, planning, decision making, and design (Howell & Cooke, 1989). As a result, a full understanding of many tasks, at a level required to intervene via training or system design, requires an examination of their cognitive underpinnings. Additionally, the growing complexity of tasks frequently surpasses the cognitive capabilities of individuals and thus, necessitates a team approach. For instance, teams play an increasingly critical role in complex military operations in which technological and information demands necessitate a multioperator environment (Salas, Cannon-Bowers, Church-Payne, & Smith-Jentsch, 1998).

Whereas the team approach is often seen as a solution to cognitively complex tasks, it also introduces an additional layer of cognitive requirements that are associated with the demands of working together effectively with others. Team members need to coordinate their activities with others who are working toward the same goal. Team tasks often call for the team to detect and recognize pertinent cues, make decisions, solve problems, remember relevant information, plan, acquire knowledge, and design solutions or products as an integrated unit. Therefore, an understanding of team cognition, or what some have called the new "social cognition" (Klimoski & Mohammed, 1994), is critical to understanding much team performance and intervening to prevent errors or improve productivity and effectiveness.

The assessment and understanding of team cognition (i.e., team mental models, team situation awareness, team decision making) requires psychometrically sound measures of the constructs that comprise team cognition. However, measures and methods targeting team cognition are sparse and fail to address some of the more interesting aspects of team cognition (Cooke, Salas, Cannon-Bowers, & Stout, 2000). In addition, to be applicable to complex multioperator military contexts, such measures need to be developed and evaluated in a task environment that is conducive to scientific rigor, yet applicable to the operational settings in which the measures will be extended. Thus, we have identified as a long-term research goal the development and evaluation of measures of team cognition in a military context. At the same time, as measures of team cognition are developed they can be used to better understand team cognition.

LONG-RANGE OBJECTIVES

The goal described above can be decomposed into the following long-range objectives:

- Identify needs and issues in the measurement of team cognition.
- Develop a military synthetic task environment that emphasizes team cognition.

- Develop new methods suited to the measurement of team cognition.
- Evaluate newly developed methods.
- Apply methods to better understand team cognition.
- Apply methods to evaluate interventions relevant to team cognition.

Thus, our long-range objectives include identification of critical issues in the measurement of team cognition, synthetic task and methodological developments, and empirical studies in the synthetic task environment to evaluate methods and to better understand team cognition. Progress during the three-year period of the effort reported here has been made along each of these fronts. Before reporting our progress in the task, methodological and empirical areas, we summarize the theoretical, empirical, and pragmatic considerations that motivated our research program. In the course of reviewing the literature reported here, we also identify needs and issues in the measurement of team cognition, our first long-range objective.

BACKGROUND

Our research program is driven by three main assumptions, each supported with a body of literature. The assumptions are: 1) Team cognition affects team performance and thus, its measurement and understanding are critical to research and applications concerning team performance in complex environments, 2) The measurement of team cognition is in its infancy and many issues and problems remain unexplored, 3) Synthetic task environments provide a useful setting for our research in that they preserve some of the best features of both laboratory and field research contexts. In this section we provide some background information supporting each assumption.

Team Cognition and Team Performance

Salas, Dickinson, Converse, and Tannenbaum (1992) define *team* as "a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/object/mission, who have each been assigned specific roles or functions to perform, and who have a limited life span of membership" (p. 126-127). Thus, teams, unlike some groups, have differentiated responsibilities and roles (Cannon-Bowers, Salas, & Converse, 1993). This division of labor is quite common in military settings and enables teams to tackle tasks too complex for any individual. Interestingly, this feature is also one that has been neglected by current measurement practices.

Team process behaviors such as communication, leadership behaviors, coordination, and planning have been linked theoretically and empirically to team performance (Foushee, 1984; Stout, Salas, & Carson, 1994; Zalesny, Salas, & Prince, 1995). Many interventions for improving team performance have targeted team process behavior (Braun, Bowers, Holmes, & Salas, 1993; Leedom & Simon, 1995; Prince, Chidester, Cannon-Bowers, & Bowers, 1992; Prince & Salas, 1993). Recently, it has become clear that other factors that are more cognitive than behavioral in nature also play a role in team performance. There has been significant theoretical work delineating cognitive constructs such as team decision making, shared mental models, and team situation awareness (Cannon-Bowers, et al., 1993; Orasanu, 1990; Stout, Cannon-Bowers, & Salas, 1996). It is assumed that with an understanding of these constructs, training and design interventions can target the cognitive underpinnings of team performance. Also, the hypothesized relation between team cognition and team performance suggests that team performance can be predicted from an assessment of team cognition, thereby circumventing the need for teams to perform in less than optimal settings (e.g., minimal training, hazardous or high-risk environments) for performance assessment.

Our research on team cognition thus far has focused on team knowledge. Parallel to research on individual expertise (e.g., Chase & Simon, 1973; Glaser & Chi, 1988), accounts of effective team performance highlight the importance of knowledge, or in this case, team knowledge. For instance, Cannon-Bowers and Salas (1997) have recently proposed a framework that integrates many aspects of team cognition in the form of teamwork competencies. They categorize competencies required for effective teamwork in terms of knowledge, skills, and attitudes that are either specific or generic to the task and specific or generic to the team. Similarly, a team's understanding of a complex and dynamic situation at any one point in time (i.e., team situation awareness) is supposedly influenced by the knowledge that the team possesses (Cooke, Stout, & Salas, 1997; Stout, et al., 1996).

Based on this theoretical work and our own observations, we have developed a framework (see Figure 1) that helps to better define team knowledge, and especially, to distinguish team knowledge as it has been traditionally measured (i.e., at the collective level) from team knowledge as it may better be measured (i.e., at the holistic level). Measures at the collective level elicit knowledge from individuals on the team and then aggregate the individual results to generate a representation of the collective knowledge of a team. Although we believe that collective knowledge should be predictive of team performance, it is also devoid of the influences of team process behaviors (e.g., communication, coordination, situation awareness), analogous to individual cognitive processes that transform the collective knowledge into effective knowledge. This effective knowledge is what we describe as the holistic level and is associated with actions and ultimately, with team performance. One of our research goals is to test this model and identify ways to measure team knowledge at the holistic level. Also note in Figure 1 that team knowledge consists of background knowledge that is long-lived in nature, as well as more dynamic and fleeting understanding that an operator has of a situation at any one point in time. Measures of team cognition have focused primarily on the former, at the expense of the latter. In general, we have found this framework useful for identifying important issues or gaps in the measurement of team knowledge.

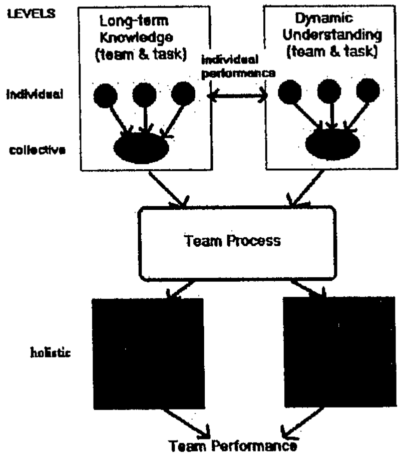


Figure 1. Framework for team knowledge.

Thus, based on the framework in Figure 1, we define *collective team knowledge* as that knowledge, both long-term and situation-specific, possessed by the aggregate of all individual team members. Second, we define *holistic team knowledge* as that knowledge, both long-term and situation-specific that is reflected in team actions and the ultimate outcome of those actions and that derives from the interaction between collective team knowledge and team process behaviors.

The Measurement of Team Knowledge

Reliable and valid measurement of constructs like team knowledge is a first, albeit nontrivial step, that presents a "road block" to advances in our understanding of team cognition. Many parallels can be drawn between the measurement of individual and team cognition, given that the primary difference is whether the measurement is directed at the team or individual level. Just as individual cognition is reflected in the behavior of the individual, team cognition is reflected in the behavior of the team. However, as discussed in the previous section, our focus on team knowledge measurement (most closely aligned with the shared mental model literature) has highlighted several areas in which measurement can be improved, including the tendency for researchers to target team cognition, by focusing on the individual level and then aggregating results.

Because research in this area is relatively new, the measures of team cognition that have been used tend to explore a small portion of the space of possible measures as is indicated in Table 1. This table classifies team knowledge measures according to type (long-term, fleeting) and the metric used to infer team knowledge from knowledge elicited from individuals. In particular, measures of intrateam knowledge similarity, and to a lesser extent an aggregate of individual team member accuracy, have been used to reflect team knowledge. These metrics do not take into account the heterogeneous background of members of many teams and consistent with the previous definition of team. Cooke, et al. (2000) suggest other metrics that take into account the knowledge responsibilities of each individual position on the team. There are other possible classification schemes not included in Table 1, such as whether the knowledge is declarative, procedural or strategic, the type of technique used to elicit the knowledge in the first place, and whether the elicitation is collective or holistic. The Xs in the table indicate the cells in which measurement work has taken place. Apparently there is much room for further development.

Table 1. Current progress in measures of team knowledge.

TYPE OF KNOWLEDGE	**METRIC**			
	Similarity	**Overall Accuracy**	**Positional Accuracy**	**Interpositional**
Long-term (shared mental models)	X	X		
Dynamic (team situation models)				

The various measurement issues relevant to team knowledge that have been identified thus far are described in detail in Cooke, et al., (2000) and are briefly summarized in the list below:

- Measures that target the holistic level, rather than the collective level, of team cognition.
- Measures of team cognition suited to teams with different roles (i.e., heterogeneous teams).
- Methods for aggregating individual data to derive collective knowledge (e.g., the social decision scheme literature (Hinz, 1999; Davis, 1973)
- Measures of team knowledge that target the more dynamic and fleeting situation models.
- Measures that target different types of team knowledge (e.g., strategic, declarative, procedural knowledge or taskwork vs. teamwork knowledge).
- The extension of a broader range of knowledge elicitation methods to the problem of eliciting team cognition.
- The streamlining of measurement methods to better automate them and embed them within the task context.
- The validation of newly developed measures.

In summary, there are many methodological gaps and a variety of issues surrounding the measurement of team cognition. Our efforts have addressed each of these issues to some degree, but we have especially focused on metrics of team knowledge for heterogeneous teams and assessment of team knowledge at the collective level.

Synthetic Task Environments

Our work has been greatly influenced by the assumption that synthetic tasks provide ideal environments for cognitive engineering research on complex tasks. We have developed an STE (Synthetic Task Environment) based on the military task of UAV (Uninhabited Air Vehicles) operations. Our research and methodological developments in team cognition have taken place in this context.

Synthetic tasks are "research tasks constructed by systematic abstraction from a corresponding real-world task" (Martin, Lyon, & Schreiber, 1998, p. 123). Performance on a synthetic task should exercise some of the same behavioral and cognitive skills associated with the real-world task. An STE provides the context for a suite of synthetic tasks. This environment offers a research platform that bridges the gap between controlled studies using artificial laboratory tasks and uncontrolled field studies on real tasks or using high-fidelity simulators.

An STE can be considered a type of simulation, but philosophically differs from traditional simulations in terms of goals and resulting design decisions. Simulations typically recreate the work environment or the equipment or systems within that environment. An STE is "task centric" in that the goal is to recreate aspects of the task to differing degrees of fidelity. Thus, an STE may not have the "look and feel" of the operational environment, but instead requires the same thoughts and behavior of the operational task. Because tasks are often situated in rich environments, STEs often include simulations of systems required to support the task. However, the focus is on abstracting task features consistent with the purpose of the planned research associated with the STE and concomitant design objectives. As a result, several very different STEs can be based on the same real task by virtue of applying distinct filters, each associated with different objectives. Such is the case with the UAV task in which a variety of STEs have been developed that focus on various cognitive skills of individuals (e.g., Gugerty, Hall, & Tirre, 1998; Martin et al., 1998) and others, such as the out UAV-STE, focusing on team cognition.

In addition, simulations often replicate the environment at the expense of the simulation's flexibility as a research tool. Researchers are limited in the degree to which they can alter or control the simulation and the measures that they can derive from it. STEs, on the other hand, typically incorporate multiple task scenarios, and often the ability to manipulate aspects of task scenarios, as well as flexibility in measurement. This increased flexibility is not so much inherent to the concept of an STE, as demanded by researchers who appreciate the benefit of trading off some aspects of fidelity for research flexibility (e.g. Fowlkes, Dwyer, Oser, & Salas, 1998; Cannon-Bowers, Burns, Salas, & Pruitt, 1998). Recently, researchers have cautioned against the use of simulations unguided by training principles or an understanding of the actual task requirements and have extolled the virtue of low-fidelity simulations that take such factors into account (Miller, Lehman, & Koedinger, 1999; Salas, Bowers, & Rhodenizer, 1998).

STEs, like high-fidelity simulations, can facilitate research in a safe and inexpensive setting and can also be used for task training and system design in support of tasks. They are also touted as providing a viable middle ground between overly artificial lab research and uncontrollable field research (Brehmer & Dorner, 1993). In many ways, STEs seem like the best of both worlds -- the laboratory and the field. Alternatively, if they fail to meet the combined objectives of experimental control and sufficient representation of the task in question, they may instead capture the worst of both worlds—poor experimental control and low fidelity.

Whereas lack of experimental control has not been a major criticism levied against STEs, lack of fidelity has. STEs have been described as *low-fidelity* simulations, as opposed to traditional equipment-centric simulations. Indeed, STEs may have low fidelity in terms of replicating the features of the equipment. The low fidelity criticism is tied to more general concerns about low face validity. However, this issue is addressed by Salas, et al. (1998), who argue that face validity may dictate acceptance by users, but not necessarily success as a training or research tool.

Perhaps more importantly, low fidelity is linked to low external validity and consequently, lack of generalizeability to the situation of interest. On the other hand, this low external validity criticism breaks down if fidelity is considered more broadly. Fidelity is generally the match between the research environment and the specific environment to which results are assumed to transfer. The match, however, can be based on a number of dimensions including the equipment and the task requirements. Thus, fidelity is not necessarily a single feature that is high-or-low for a particular simulation, but rather a multidimensional feature that can ultimately result in contexts of mixed fidelity. That is, a simulation may be faithful to the equipment, but not to the task requirements. In light of the issue of external validity, it is important to determine the dimensions of the transfer situation that are relevant to the research questions to be generalized. A mixed fidelity simulation may have adequate external validity, and thus generalizeability to the field of practice, if it is faithful to the relevant dimensions of the field of practice. Determining external validity then becomes a question of accurately identifying the relevant dimensions in the field of practice for the research questions. Generalizing results to other fields of practice amounts to identifying similar features along the same relevant dimensions in other fields of practice. It can then be assumed that the match is sufficient for research results to generalize to this environment. This enterprise of identifying and matching the features and dimensions among different work environments amounts to a theory of tasks or work environments.

Under this multidimensional view of fidelity, the labeling of traditional simulations as *high fidelity*, and of STEs as *low-fidelity*, makes little sense. Instead, STEs are typically high fidelity with respect to the task and low-fidelity with respect to the equipment. Traditional simulations may more often be associated with the opposite pattern. External validity cannot be determined independent of the research questions.

Research on cognitive aspects of a complex task such as decision making under stress, may best be addressed in an context that preserves the features of the task at the expense of the fidelity of the equipment. Alternatively, research directed at uncovering reasons for operational errors associated with a piece of equipment may generalize only in a context that faithfully replicates the equipment, perhaps at the expense of simplifying the overall task. The question of the external validity and extent of generalizeability of both traditional simulations and STEs needs to be addressed for each test-bed in the context of each research question.

One of the first steps under the effort reported here was to design and develop an STE that would provide a rich test-bed for methodological developments and empirical investigations in the area of team cognition.

Summary: Background

The literature on team performance and cognition has pointed to a pressing need for better measures of team cognition, as well as for the evaluation of those measures. We have identified several issues relevant to the measurement of cognition at the team level. Synthetic task methodology provides an ideal environment to develop and test measures and to investigate team cognition. Synthetic tasks preserve much of the complexity and richness of the field-of-practice, yet afford greater experimental control than field studies.

PROGRESS UNDER THIS EFFORT

In the previous section we described the needs and issues in the measurement of team cognition that we identified in accord with our first long-range objective. In this section we summarize our progress toward each of the five remaining long-range objectives that include:

- Develop a military synthetic task environment that emphasizes team cognition.
- Develop new methods suited to the measurement of team cognition.
- Evaluate newly developed methods.
- Apply methods to better understand team cognition.
- Apply methods to evaluate interventions relevant to team cognition.

A section is devoted to progress toward each of the first two objectives and progress toward the final three objectives is described in a section on empirical studies.

DEVELOPMENT OF AN UNINHABITED AIR VEHICLE SYNTHETIC TASK ENVIRONMENT

UAV operations, a task environment of increasing interest in both military and civilian aviation arenas, was selected as the field of practice, with the Air Force's Predator UAV operations serving as the specific work environment. The selection of the UAV work environment was also based on a need for a team STE, which was satisfied by the fact that this task requires team performance and is associated with interesting team issues involving rank differences, the team's interaction with automation, and the presence of hierarchical and distributed teams. Also, the UAV operations task involved team planning, decision making, and situation awareness and thus, provided a suitable test bed for research on team cognition.

Although a final test of the external validity of our STE awaits field trials in which the findings and measures derived in the STE context are applied to the field of practice, the design decisions that were made along the way took external validity and generalizeability into account. In the following sections we describe the steps involved in developing the UAV-STE including the identification of design constraints, the abstraction of task features, prototype development, and iterative testing and re-design.

Design Constraints

Before the abstraction of task dimensions and features took place, it was necessary to acquire a thorough understanding of the field-of-practice and the research objectives. Together, these constrained the abstraction process. In addition, other constraints also surfaced, such as the expertise of the intended participant population and various technological constraints in replicating task features in the laboratory.

Information from the field of practice. Paramount to the development of an STE is an understanding of the actual task, based on documentation, interviews with experts, examination of other STEs for that work environment, and behavioral and cognitive task analyses. UAVs come in a number of varieties, but our focus was on the Air Force Predator, the first operational Air Force UAV, as well as one for which we had access to training data, operational specifications, and subject matter experts. In particular, we relied heavily on a cognitive task analysis (CTA) done on UAV operations at Indian Springs, NV (Gugerty, DeBoom, Walker, & Burns, 1999), information from actual Bosnia operations, an existing UAV-STE at Williams AFB in Mesa AZ (Martin, et al., 1998), and discussions with various investigators involved with these projects. Other information was obtained from various Internet sites and unpublished reports, especially training documentation for the Predator UAV. From this information we began to develop an understanding of the UAV work environment.

By way of brief summary, the Predator UAV is controlled by operators in a GCS (ground control station) who communicate with other groups concerning issues of data interpretation and airspace deconfliction. The major team members within the GCS include the AVO (Air Vehicle Operator) who operates the UAV, the PLO (Payload Operator) who operates the sensors, and the DEMPC (Data Exploitation, Mission Planning, and Communications Operator) who is responsible for mission planning. These individuals work together to accomplish the goal of navigating the UAV to a position to take reconnaissance photos of designated targets. Individual team members have access to information about the UAV flight system, sensor equipment, and the surrounding environment, by way of computer displays, hard copies, and communication channels.

Information from research objectives. The research objectives also constrained the STE. That is, different features from the field of practice were abstracted, ultimately resulting in a UAV-STE distinct from one developed under a different set of research objectives. For instance, an STE that is developed for training UAV operators would appear quite different from one designed to develop models of team performance or one designed to test the effects of automation on team performance. In our case, the STE was designed to serve as a flexible task environment for the development and evaluation of measures of team cognition. This overall goal resulted in the following three objectives: 1) the STE should facilitate the measurement of team cognition, 2) the STE should provide a realistic task environment, and 3) the STE should provide an experimenter-friendly research test-bed. Each of these is discussed in depth below.

1) The STE should facilitate the measurement of team cognition. The *team* component of this objective means that the tasks and scenarios generated in the context of the STE should require a team of individuals. Based on the definition of *team* discussed previously, it is important that each team member has a distinct role, and that without interdependence among individuals, the task cannot be completed.

   The *cognition* component of this objective requires that the task environment pose cognitive demands both at the individual and team levels. For example, individuals may need to apply their own knowledge, and also share knowledge with fellow team members. In other cases, the individual situation assessments by team members may feed into the assessment of that situation by the team as a whole. In particular, the intended focus on the measurement of team knowledge requires that the STE support task that are knowledge- and information-intensive, with opportunities for variations in knowledge distribution across team members.

2) The STE should provide a realistic task environment. In general, this same objective motivates all STEs. It is based on a need to study behavior and cognition at a level at which it occurs in complex real-world tasks, yet to do so in the controlled environment of the laboratory. Specifically, for our purpose, it is important that the cognitive measures be developed and validated in a context that is relevant to real-world team tasks, and at the very least to the team tasks of the operational Predator UAV environment. This goal will be achieved to the extent that the STE exercises the same kinds of cognitive and team skills as the task in the field of practice.

   Realism, like fidelity, is multidimensional and the dimensions of importance will vary with the research objectives. In our case, the STE needs to be realistic at the level of team cognition and knowledge requirements. Under a different set of research objectives, realism in terms of face validity to actual operators or in terms of motor control of system devices may be more important.

3) The STE should provide an experimenter-friendly research test-bed. One of the purposes of collecting information on existing STEs and from investigators who use them was to situate the STE in a relatively unique niche, partially by benefiting from the lessons learned by others. Specifically, we uncovered a voiced need to make the STE more compatible with experimenter needs. In a sense, this objective involved moving away from a focus on simulation and toward a focus on research.

   For instance, the STE should be flexible in supporting synthetic task variations of the type that may be required to manipulate a variable of interest or create a task scenario with slightly different cognitive requirements. It should allow, for example, rapid manipulation of task complexity, workload, cognitive demands, team process constraints (e.g., intrateam communication), as well as details of task scenarios. Further, it should allow the experimenter to intervene in the scenarios in order to insert measurement probes or to alter the task on-line (e.g., respond to varying levels of performance by increasing task difficulty during a mission). Similarly, the STE should be flexible in supporting a variety of cognitive and performance measures. Specific measurement capabilities should include embedded performance measures, logs of and post processing routines for computer and communication events, and rapid data access for immediate analysis.

Information from other constraints. There were other constraints associated with the limitations of the intended research environment. Specifically, there were constraints associated with the fact that the STE was to be embedded as one of several STEs within NMSU's CERTT (Cognitive Engineering Research on Team Tasks) Laboratory, a unique and flexible facility designed to host a variety of synthetic tasks for teams, especially those in a military context (Cooke & Shope 1998). The CERTT Lab was funded by a 1997 DURIP grant (F49620-97-1-0149). The CERTT hardware and software configurations provide a number of benefits in terms of overall flexibility and cognitive measurement. The hardware consoles have both computer and video monitors and general communications hardware interfaces that are appropriate and realistic to a wide variety of STEs. However, this same flexibility also present some constraints. For example, the consoles have a general military-equipment look to them without using any hardware features uniquely associated with an actual UAV operator console (see Figure 2). Instead, we migrated STE-specific functions to software displays.

Of primary importance to the design of the lab is that the facility provides maximum flexibility for research needs associated with data collection, measurement, and control and manipulation of various factors in STEs. For instance, CERTT hardware consists of four participant consoles and an experimenter control workstation that can be arranged to simulate distributed or co-located team contexts.

Each of the participant consoles (see Figure 2) consists of two NT-based workstations and monitors, a video monitor, a joystick, a keyboard and mouse, a headset and intercom, along with a variety of other lights, power switches, audio mixers, and indicators. A network of ceiling-mounted cameras permits the experimenter to view and record each of the four participants via a video monitor and video recorder contained in the experimenter console. Careful planning went into the communication module of the workstations. The headsets and intercoms provide direct experimenter control of various communication pathways among team members. A digital audio switching matrix allows a very flexible communications system. The experimenter can disable specific communication links. In addition, audio data is recorded in sync with other data records (i.e., time-stamped) and identifies not only time of communication, but speaker and listener's identities as well. The communication module also creates auditory isolation of team members when necessary. The four participant workstations and the experimenter station are connected locally using a 100BaseT local area network. This LAN is used by the synthetic task software for event synchronization and real-time data transfers.

Figure 2. Participant consoles in NMSU's CERTT Lab.

Additionally, the CERTT lab has Internet connectivity. A separate computer, designed to serve as a gateway router was recently installed in the laboratory. This router will take the simulation data flowing within the CERTT LAN via DDE packets, and reformat them for transfer over the Internet. We are using VR-LINK as a distributed interactive simulation (DIS) tool. At some remote location, the packets are reconstructed into local DDE packets; this allows a remote CERTT console to participate in future CERTT experiments. We call this a distributed STE.

The UAV-STE was the first synthetic task designed for the CERTT Laboratory. Many of the CERTT Lab's hardware design, development, and construction phases were done in parallel with the UAV-STE development. This permitted close developmental cooperation between this particular STE and the overall CERTT Lab design. This was a valuable feature not likely available to future STE development exercises. However, the overall lab development goal was one that preserved a high degree of flexibility and adaptivity to meet the experimental goals of a wide range of STEs.

In addition, the CERTT Laboratory supports team tasks using up to four isolated participants and relies on a participant population of college students. Thus, the necessary STE skills and knowledge needed to be learnable with little or no background knowledge and without extensive training and practice. Time was also a factor here in that good team data should be obtainable within the time frame of a semester. This also puts limits on the amount of time that a team can spend in a training and skill acquisition phase.

Abstraction of Features

The next step involved taking all of the information gathered from the field of practice, the research objectives, and other constraints and using it to abstract aspects of the task suitable for the STE. We identified those aspects of the actual task that we planned to emphasize (or maybe even exaggerate) in our STE. Due to our research objectives, these were features of the task that centered on team cognition.

1) In the field-of-practice, background knowledge and information relevant to UAV operations are *distributed* across team members. In other words, some UAV knowledge is uniquely associated with individual team members. For instance, the AVO has knowledge of the UAV's flight controls and equipment, whereas the PLO has knowledge of the sensor equipment. In a similar way, each team member is provided with information during the mission that is relevant to his or her role. This feature of distributed knowledge and information across a team was one that seemed relevant to team cognition and one that we chose to focus on in the STE version of the task.

2) Although information is distributed across team members in the field of practice, there are significant information interdependencies such that *knowledge and information sharing* among team members is required. For instance, the DEMPC determines the targets and provides navigation information to the AVO and PLO. The AVO, in turn, provides information relevant to the UAV that assists the DEMPC with route planning. The PLO communicates to other team members the navigation requirements for achieving target objectives. Thus, this feature of the task in which team members were required to share knowledge and information, not only filled the team constraint, but also the cognitive one, and was retained in the STE version.

3) UAV operations involve extensive *planning* on the part of the team in regard to the UAV's route. This planning occurs under multiple constraints. For instance, in order to plan the route, the PLO considers ground distance, altitude, depression angle, and slant range relevant to sensor operations; the AVO considers spatial layout and speed, and the DEMPC considers features of the bigger picture (i.e., restricted areas, ROZ boxes (Restricted Operating Zones), terrain height, weather, and speed). The planning process is a highly cognitive one, and in the UAV task, requires cognition also at the team level, because the various plans of individuals

need to be integrated into a single plan. Thus, planning was retained as an integral component of the UAV-STE.

4) Finally, due to the dynamic nature of the UAV operations environment (i.e., the presence of ad hoc targets, weather changes, UAV malfunctions, cloud cover) there is a need for *dynamic replanning* by individuals and the team as a whole. Team situation awareness is also required to plan and replan in an effective and efficient manner. Therefore we decided that the same kinds of dynamic events that occur in the field of practice needed to be preserved in the UAV-STE in order to provide the opportunity for dynamic replanning.

Prototype Development

The next step in our design process consisted of transforming the ideas, constraints, and abstracted features generated in the first two steps into a prototype STE. We chose to use paper mock-ups, simple drawing software for screen designs, and extensive functional specifications. The difficulties that we encountered at this step were related to the fact that our design called for three independent, yet interconnected participant systems, as well as an experimenter system that would serve to monitor, manipulate, and measure aspects of the STE and even at times play a participatory role in the STE. Thus, a major part of our prototyping efforts involved determining interdependencies among the various systems.

We started this part of the design process by using the abstracted features and various design constraints to determine the minimum set of functions that each team member and experimenter would be required to perform. For instance, the PLO needed to monitor and adjust camera settings. We limited these to focus, shutter speed, aperture, zoom, and camera type. Each step of the way, we determined functionality with the goal of preserving task realism and especially functions that led to interdependencies (e.g., the shutter speed was tied to the airspeed of the UAV which was controlled by the PLO so this function was viewed as critical), but at the same time keeping in mind the background knowledge and training limitations of our participant population.

Once we settled on functions for each of the team roles we began the task of designing the interface. The CERTT Lab participant workstations each consist of two computer screens and so we decided to use one for primary control functions and the other for more secondary monitoring functions (i.e., warnings and alarms) of each team member. We made an intentional leap from realism where the interface is concerned in the interest of minimizing training time. The Predator interface is highly complex and can require a lengthy acquisition period (up to a year) for Air Force personnel to reach asymptotic performance. As our objectives centered on performance of a team with members already versed in their individual interface operation, the replication of the Predator operations interface was not only unnecessary (as functionality of the interface was preserved), but would be a hindrance to collecting data on teamwork and team cognition.

Next, we developed a list of software data structures needed to support the prototype design. These structures were further broken down into dynamic variables used to exchange information between the various computers on the network, startup variables, and data stores used to record performance information. An integrated system of data files was also defined to permit the experimenters to change startup parameters, record operational raw data, and to archive performance statistics derived from the various real-time measures. The initial DDE (network communications) architecture was considered at this time as well.

Because the design process for the UAV-STE was concurrent with the design of the hardware in the CERTT Lab, we were able to mock up the STE along with the supporting hardware. This aided the hardware developers by providing an actual example of an STE. The STE and consoles were simultaneously mocked-up using paper and cardboard prototypes. Figures 3 shows an example of one of the prototypes that we used.

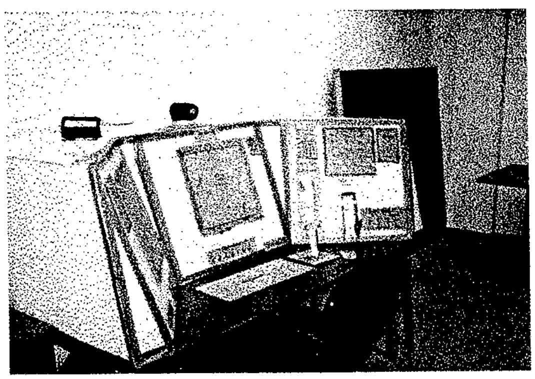


Figure 3. Paper mockup of STE and subject console.

Functional specifications were drawn up in detail and provided information regarding interface displays and controls, measurement requirements, functional properties of subsystems, and information flow between start-up files and other subsystems. Our resulting prototype consisted of seven interconnected systems (two for each of three team members and one for the experimenter), and for each system, a representation of the screen and detailed functional specifications. The final version of the functional specifications was also key to the design of our training program for the UAV-STE.

Implementation and Iterative Design

The next step involved presenting the functional specifications to the programmers who reviewed them and provided feedback on technological constraints and software architecture that would impact the basic design. This process began a series of discussions between STE designers and programmers and concomitant feedback-redesign iterations. The start of the implementation process further prompted additional feedback and redesign iterations that continued through the remainder of the UAV-STE development. Reliability and straight-forwardness were two important design goals as well; not only for the developmental phase, but, for hardware/software maintenance and ease of future upgrades and or modifications.

The software architecture for the CERTT Laboratory is based on a Windows-NT local area network. Each computer is running a local application specific to the display and function assigned to any particular screen. The experimenter computer serves as the master control unit for the entire suite of applications. Each application is started and placed in a standby mode on startup. When all applications are standing by, the experimenter can start the simulation simultaneously. The simulation is operated in a pseudo-real-time mode using DDE data packets broadcast over the local area network. Particular applications make use of only those packets relevant to their function.

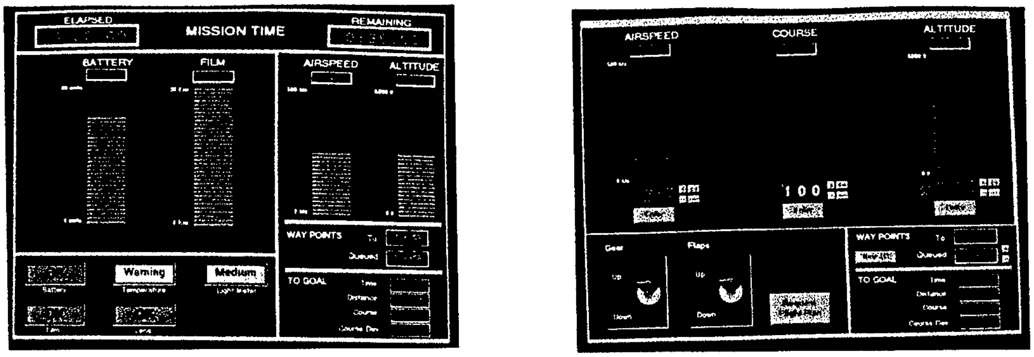
The experimental results of any individual mission are stored in data files located on the experimenter's computer. This allows ready archiving and retrieval of the performance data at the end of a mission. The mission setup parameters are also stored in this fashion as well and are used by the various applications while initializing.

The Resulting UAV-STE

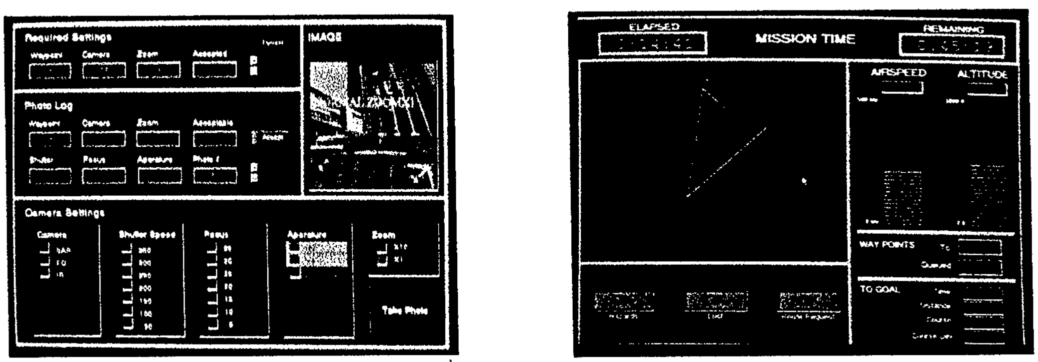
In the resulting UAV-STE three team members work together to control and navigate the UAV and to take photographs of designated targets. The AVO is generally responsible for controlling the UAV's heading, altitude, and airspeed, as well as monitoring landing gear, flaps, and fuel consumption. The PLO takes photographs of designated targets under camera settings that are specified on the basis of destination waypoints and the current UAV speed and altitude that are broadcast from the AVO's computer to the other participants. The PLO also monitors the camera equipment such as battery, film, temperature, and lens. The DEMPC is responsible for mission planning and thus, has information that provides the "big picture" of the mission such as the world map that indicates all of the waypoints available. The DEMPC is presented with an initial route plan as the experiment begins, but has the capabilities to edit this plan as the experiment progresses. In general, the primary computer display at each console provides the information and controls central to the individual's task and the secondary screen to the right contains system status displays that need to be monitored and action taken in the event of a warning or alarm. These secondary screens also provide the capability to increase workload by manipulating alarm rate. The six screens associated with the three team positions are displayed in Figures 4a-4f.

Other characteristics of the UAV-STE support the objectives discussed previously. Flexibility is achieved through task parameters, a waypoint library, and embedded performance and communication measures described in more detail in the next section.

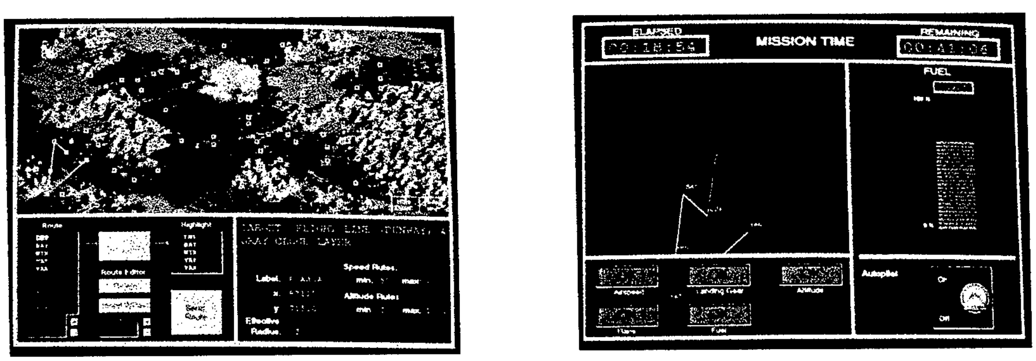
Also, the interface requirements for the STE were based on the functions required to carry out the abstracted UAV tasks. Whereas the details of the displays and controls are somewhat dissimilar from the UAV interface in the field of practice, the new controls and displays support the new tasks and reduce the time needed to acquire proficiency.



Figures 4a and 4b. Screens from the AVO participant console



Figures 4c and 4d. Screens from the PLO participant console



Figures 4e and 4f. Screens from the DEMPC participant console

Specifically, we had learned that interface complexities of the actual system resulted in extensive training requirements (up to a year). Our modified UAV-STE interface, which preserves much of the functionality of the Predator interface, requires only a one-hour tutorial, followed by a 30-minute practice session to get team members to the point at which they can accomplish their individual tasks and begin to work together. Thus, this change in interface from the actual task reduces training time for the student participants and facilitates focus on the aspects of the task that involve team cognition.

As a result of the abstraction process, the UAV-STE, in comparison to the UAV task in the field of practice, exaggerates team cognition and interaction. Some aspects of the Predator UAV operations involve minimal or no team interaction and these have been eliminated or minimized in importance. For example, pre-flight planning that is DEMPC-focused and take-off and landing procedures that are AVO-exclusive have been eliminated in the UAV-STE. In addition, the UAV-STE retains the distributed nature of the task information and in fact, increases the compartmentalization of this information so that there are more cases in which information is uniquely associated with a single individual, thereby increasing interdependencies. To illustrate, in the STE, unlike the field of practice, the DEMPC is solely responsible for route planning from target-to-target and has sole access to a "big picture" map. In a similar way, information interdependence and the need for team coordination and knowledge sharing have been exaggerated by adding or modifying some rules that depend on information integration across two or three team members. For instance, the AVO and PLO both have altitude constraints concerning the UAV flight path and sensor operation respectively. Both have to work together to simultaneously satisfy these constraints.

Finally, the UAV-STE provides an ideal environment for the development of knowledge measures, because it offers experimental control over the information presented to the team members, the information acquired in training (i.e., knowledge), and team experience. Thus, for the purposes of measure development, a team can be constructed in which knowledge and information are completely and nonredundantly distributed across team members or in which all team members are trained on, and presented with, identical information. The ability to characterize, with any certainty, a team's experience and knowledge base is a luxury unique to the synthetic task environment. Characterizing tasks and teams in regard to knowledge distribution or experience in the field is challenging at best, and artificial lab tasks fail to provide the information richness needed for work of this kind. With this kind of information in hand, newly developed knowledge measures can be applied to team tasks in which the answers (i.e., content and distribution of knowledge) are known. A good measure should be capable of identifying those known features and distinguishing teams that have different knowledge profiles.

Summary: Development of a UAV-STE

Under this effort we have designed and developed a synthetic task environment for teams based on actual operations of the Predator UAV. The design of the UAV-STE took into account constraints from the field-of-practice, research objectives, and the CERTT Lab's facilities in the abstraction of features for the synthetic version of the task. Development of functional specifications and prototypes and iterative evaluation and redesign helped to create a rich test bed for the development and testing of team cognition measures and for the empirical investigation of team cognition.

METHODOLOGICAL DEVELOPMENTS

In accord with the long-term objective of "develop new methods suited to the measurement of team cognition" we developed a number of software tools for facilitating experimental control, new measurement methods, adaptations of existing measurement methods, software tools for automating measurement, and various quantitative metrics for assessing team knowledge and performance.

This methodological development effort is ongoing and began in the first year of this effort in parallel with the development of the UAV-STE. In particular, we found it important to consider the experimental control and measurement issues as the task was being implemented so that data from the task was captured in a form suitable for the measures. In order to pre-test some of these methods, we pilot-tested very early versions on small numbers of participants. In addition we applied some of the later versions to experiments completed under a different effort using a lower-tech synthetic task environment (helicopter search-and-rescue mission; Cooke, Kiekel, Salas, Stout, Bowers, & Cannon-Bowers (2001)). Results from these pilot tests and studies have provided rich data for revision of methods and have guided much of the development work discussed below.

Experimental Control Methods and Tools

Several software and hardware tools were developed to facilitate research from the experimenter point-of-view. In particular, it was important that the details of the scenario be easy to modify by an experimenter and that the experimenter's role as an observer was supported.

Task parameters. The scenario flexibility was achieved through task parameters (e.g., mission time, fuel consumption rate, malfunction frequency) that can be easily changed by the experimenter via a series of set-up files. The files are all text-based and allow easy manipulation using a simple text editor such as *Notepad.* For instance, the cognitive workload of the scenario can be manipulated by changing the frequency with which timed warnings and alarms occur.

Waypoint library. Spatial characteristics of the scenario (terrain features, weather obstructions, enemy activity, target priority, ROZ box location) can be modified through a *waypoint library* that defines the scenario in terms of a world of waypoints, each with varying features. The waypoint library is stored as text, but has a Microsoft Access database interface (see Figure 5). Through this graphical interface an experimenter can navigate to a page of information associated with a particular waypoint and review and modify features associated with that waypoint.

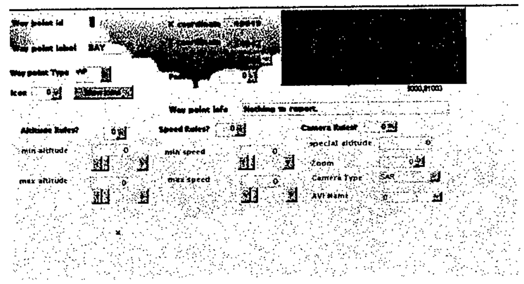


Figure 5. Waypoint Library user interface

Experimenter console. In our UAV-STE studies, there are two experimenters present at all times working together to carry out the various experimenter functions. Through the hardware and software associated with the CERTT Lab's experimenter console (see Figure 6), the experimenters can control and monitor the UAV-STE, observe team behavior, initiate data collection tasks, and archive collected data. The task software is set-up, initiated, and terminated from the experimenter console. Because the experimenter console is physically located in a separate room of the CERTT Lab, apart from the participant consoles, observation of team behavior is accomplished by using a variety of audio, video, and computer screen displays of the team during a UAV-STE

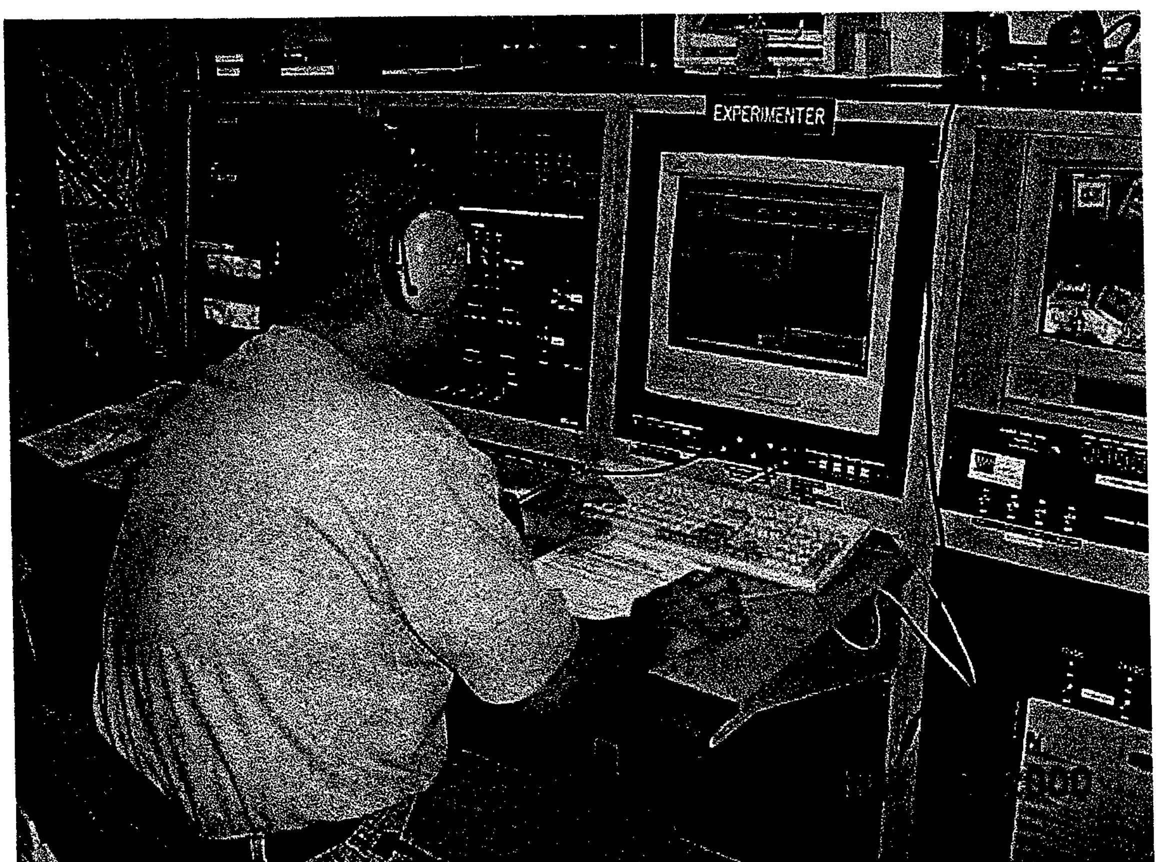


Figure 6. Experimenter at the experimenter console

mission. General observations can be entered, time-stamped, and recorded in a custom-built experimenter observation log on the console's computer screen. After the experiment, various data records are labeled and saved on removable storage media. The experimenters can also act as players in the UAV-STE. For example, an experimenter can represent a higher authority (e.g., intelligence) and can periodically call-in new waypoints or request information that is used to test situation awareness. The experimenter console can be used to interfere with communications by blocking specific communications links or to disrupt a link by injecting noise.

Training Modules and Tests. Although the interface changes implemented in the synthetic version of the UAV operations task result in fast task acquisition (1.5 hours to master the individual skills), training is nonetheless required. Powerpoint Modules were created to present 1) general information regarding the task and the three roles, 2) specific information regarding the interface controls and displays, and 3) specific information regarding facts and rules pertinent to each position. The first Powerpoint training module was the same for all three task positions, but the second and third modules were specific to each position. For the second study, some additional material regarding the other positions (i.e., in the Shared-Knowledge condition) or repeated presentation of existing material (i.e., in the control condition) was also included. Trainees could read through the module at their own pace, jumping back to previous slides if desired.

Following each of the three training modules a ten-item multiple choice test was administered on the computer. The ten items needed to be answered correctly in order for the trainee to move on to the next training module. If an error was made, the trainee was to go back to the Powerpoint module to find the correct answer. No feedback was given regarding the correct answer. The trainee then took the same test again, correcting any previous incorrect answers. If one or more errors was made a second time, an experimenter intervened to identify the missed questions and help the trainee understand the correct response and again take the test. The experimenter intervened so that one or more trainees would not fall behind the others in the training part of the study.

Measurement Methods and Tools

Performance score. Although the focus of the CERTT Lab and the UAV-STE studies is to measure and understand team cognition, it was critical to apply valid measures of team performance and process behavior to the task, because these measures would serve to ultimately validate the newer measures of team cognition. In this regard, outcome measures of team performance were given particular attention.

Various indices of individual and team performance, embedded in the UAV-STE task were recorded and summarized at the end of each mission. These include amount of film and fuel used, number and type of photograph errors, route deviations, time spent in warning and alarm states, and specific waypoints visited.

Scoring software was generated that created a composite team score, as well as individual scores, from these performance indices. The software also summarized various elements of the composite score. The composite score used in Study 1, subtracted points from a total of 1000 for unphotographed targets, unvisited critical waypoints (e.g., ROZ box entry, exit points, targets), seconds in an alarm state, and fuel and film used. The score used in the second study included these indices plus penalties for seconds in a warning state, as well as for violations in route rules (i.e., priority targets not visited first in a ROZ box, ROZ entry and exits not visited).

Several generations of performance scoring software were developed in order to include the additional penalty indices and to improve the user interface. The final version of the scoring software allows the experimenter to alter the scoring formula, thereby facilitating experimentation with different composite performance scores. This is accomplished by weighing the various performance indices by varying amounts. For example, fuel usage could be given a lower weight relative to ROZ violations so that the composite score would reflect the importance of route planning over fuel conservation. Figure 7 shows an example of several screens of the scoring software.

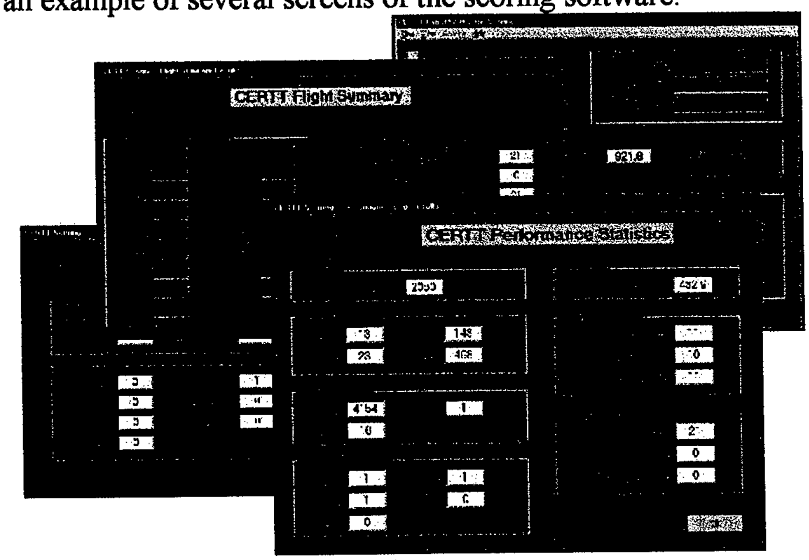


Figure 7. Several scoring screen examples

Communication recording and analyses. Communications recording and data analysis is an important component of the CERTT Lab and provides an index of team process behavior in the context of the UAV-STE. Team participants (up to four plus an experimenter) communicate with one another over military aviation headsets with microphones. The noise isolating properties of the headphones along with the use of noise-canceling microphones makes it nearly impossible to hear extraneous noise. The audio isolation, along with the physical shielding provided in the design of the consoles, result in the participants becoming rapidly immersed in the task. Furthermore, the audio and physical isolation provides a strong incentive for all participants to communicate through the headsets.

The digital communications system is quite advanced and highly flexible. The system design allows a talker (who initiates a communications episode) to select a listener or a group of listeners. The talker initiates communications by pushing and holding down a push-to-talk (PTT) button. All communications are designed as simplex. In other words, when A is talking to B, B cannot automatically talk to A without B first selecting A as a listener and pushing B's PTT button. We treat A talking to B as a distinct event as compared to B talking to A. The system is additive for incoming communications traffic. For example if A is talking to B and C also begins talking to B, then B hears a mixed audio signal made up of A and C. The system allows for simultaneous networked communications. For example A can talk to B and C while at the same time D is talking to A and B.

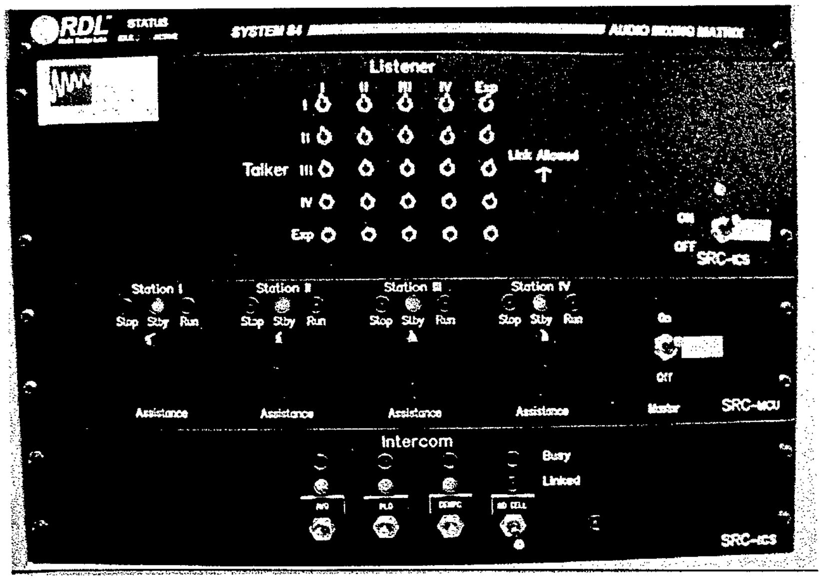


Figure 8. Experimenter communications control modules.

Additional features of the system allow the participant to be listening to the computer audio output over the headsets including alarms, warnings, and other audio clues built into the task scenario. Communications traffic, when present, over-rides the computer audio. The computer audio will return when the communications to a participant ceases. The system allows audio noise including static, random noise, recorded distracting noise such as jet engine sounds, and non-relevant communications to be added to particular communications link. This has the effect of making some links less attractive to use than others. Furthermore, we can completely disable any individual link in the system with the throw of a switch (Figure 8). For example we can allow C to talk to D but disallow D from talking to C.

The headset microphone output for each participant is recorded continuously, even when the PTT button is not depressed. This allows spurious individual utterances and talk-aloud statements to be recorded in addition to the intentional communications episodes. We record this microphone data on a VHS recorder as well as an 8-channel digital audio tape deck using a 48 kHz sampling rate. These latter recordings are of digital production quality. We also generate a mixed composite of all 8 channels and record this on the audio track of our video camera recorder. Two or more conversations or a message with an ambiguous sender can be disambiguated using this 8-track feature.

Most relevant to the communication flow problem, however, is the CERTT Lab's capability to record the precise timing and duration of messages from specific senders to specific receivers. In addition to the digital nature of the entire communications system, this capability is the result of two other features:

1) a push-to-talk button that must be depressed for the duration in which a message is being sent in order to be heard by the listener, and

2) a bank of intercom toggle switches that enable individual talkers to specify one or more listeners of the message.

The CERTT custom communication log software (Figure 9) samples the positions of the switches and the push-to-talk buttons at an experimenter-selectable sample rate. We typically use a 1 Hz sample rate (i.e., once a second). An $n \times n$ matrix is used to represents the state of the communication network made up of n team members. The rows represent senders and the columns, receivers. At each sample interval, we record a snap-shot of this matrix. This matrix represents all possible states of the communications network, including asymmetric or directed communications. The link-disable feature described above is reflected in this matrix by certain elements always being in the "off" state.

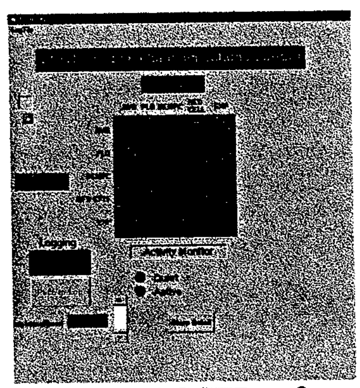


Figure 9. COM-LOGGER software screen

A separate post-processing module (Figure 10) is used to analyze the communications data and provides statistics such as number of communications episodes, the average length of each episode, etc. The advanced analyses of these communication flow data and the analysis of the content of communications, using automated methods such as Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998) are the subjects of additional ONR-sponsored research in the lab.

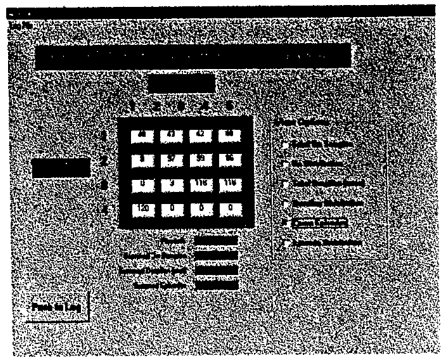


Figure 10. COM-LOGGER post-processing screen

Other process measures. Communication is one of the richest manifestations of team process behavior, however it was important that we also capture other team process behaviors such as leadership behaviors, coordination, and planning. In earlier studies in the context of the helicopter search-and-rescue mission, we attempted to measure team process by having trained judges view video tapes of team interactions and rate them along various dimensions of process (e.g., planning and decision making, communication, etc.). We learned prior to this that it was important to separate process from outcome and that the judgments of each should be kept as independent as possible.

After having little success with the process ratings method and after several discussions with our consultant, Clint Bowers, we decided to approach process behaviors more specifically. That is, we identified very specific process behaviors that a team may or may not exhibit at a particular point during the mission. We call the "particular point" the trigger event. This is generally a point after which a significant event has occurred (e.g., and ad hoc target is called in). Then, trained judges determine at these trigger events whether the specified behavior has occurred (e.g., the DEMPC communicates the change in target to the AVO). We used this type of process measurement in both empirical studies. Questions were modified for the second study, based on how well they worked in the first study. In addition, two judges were used in the first study and because agreement was so good, we reverted to one judge for the second study. Process questions asked of the judge or judges are presented in Appendix A.

We view the adequate measurement of event-triggered process behaviors in the beginning of a research program like ours, a challenge. In order to identify suitable trigger events and team process behaviors stemming from those events that discriminate teams, much insight to the typical team behavior in the task is needed. Such insight may not be possible until several iterations of experimental observations and process measurement refining.

Situation awareness measurement. Situation awareness at an individual or team level can be considered process if one considers the situation assessment behavior that leads to situation awareness. It can also be considered knowledge if one considers the situation model that is generated through the process of situation assessment. In concert with the focus of our research on team cognition and knowledge, we assess situation awareness in terms of the resulting situation model. Thus, situation awareness is measured in the UAV-STE task in terms of knowledge that is dynamic and fleeting in conjunction with the situation (i.e., the right hand side of Figure 1). We assess this knowledge by asking queries of individual team members during the mission. Our queries are more like SPAM queries (Durso, Hackworth, Truitt, Crutchfield, Nikolic, & Manning, 1998), than SAGAT (Endsley, 1995) queries in that the information displayed to the participant remains available through the questioning. The queries cover information about the task and therefore assess the team situation model regarding taskwork, as opposed to teamwork. We further assume that team situation awareness is the aggregate of the situation awareness of the individual team members. Thus, we assess situation awareness at the level of collective knowledge. Our queries are asked of individuals by the experimenters at randomly determined times during each mission. The list of queries is presented in Appendix B.

Taskwork ratings and taskwork consensus ratings. Longer-term knowledge regarding the task was measured using a pairwise relatedness rating task. Rating data like these have been used often to assess the structure of conceptual knowledge (e.g., Schvaneveldt, Durso, Goldsmith, Breen, Cooke, Tucker, & DeMaio, 1985; Cooke & Schvaneveldt, 1988). In the rating collection task, pairs of task-related concepts are presented one-at-a-time and in a random order. Order of items in the pair is counterbalanced across participants. The rating data are typically analyzed using multivariate scaling techniques such as multidimensional scaling, cluster analysis, or Pathfinder network scaling,

Custom software was created in order to collect the ratings from individual team members (Figure 11), as well as from the group. Our group or taskwork consensus rating task was a new development directed at measuring team knowledge at the holistic level (see Figure 1). In previous applications of the rating method to team knowledge measurement, team knowledge has been assessed by aggregating the individual ratings (Cooke, et al., 2000). In our studies, we investigate how the aggregate of the individual

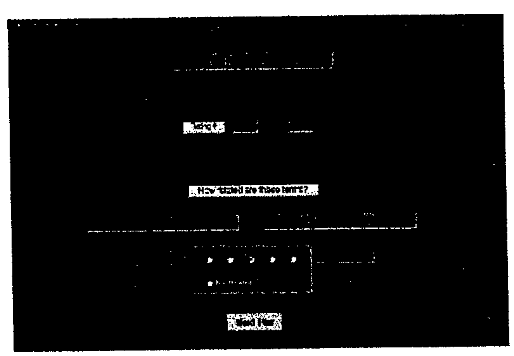


Figure 11. Single Ratings software screen

ratings (i.e., collective taskwork knowledge) compare to the consensus ratings (i.e., collective taskwork knowledge) in predicting performance and how the individual ratings map on to the consensus ratings.

Consensus ratings are collected after individual ratings have been collected. Team members are seated at their individual consoles and communicate over headsets. Pairs are randomly presented one-at-a-time to the team. Each team member sees the pair in question, along with the ratings given to that pair individually in prior sessions by all three team members. The team members must come to consensus on a rating (each entering the same rating) in order to move to the next pair.

In Studies 1 and 2 of this effort, the following 11 pairs of items were presented in both individual and consensus taskwork rating tasks: altitude, focus, zoom, effective radius, ROZ entry, target, airspeed, shutter speed, fuel, mission time, photos.

Taskwork and teamwork questionnaires. Questionnaires were also designed to measure taskwork and teamwork knowledge. In these cases the knowledge was elicited at the individual level, and aggregated in accord with the collective view of team knowledge. Taskwork questions were based on a structural decomposition of the UAV-STE task. Not only did this questionnaire prove difficult to score, but it provided little informative data and thus was not used in Study 2. The teamwork questionnaire focused on the communication links between individuals and the type of information conveyed along these links. It was more successful than the taskwork questionnaire and was revised for the second study. These questionnaires and scoring information can be found in Appendix C.

Debriefing Questionnaire. The debriefing questionnaire contained demographic questions as well as questions regarding how well the participants liked their team and the task (see Appendix F). Software was written to present this questionnaire and collect responses on the computer screen (Figure 12). Data are summarized in a spreadsheet, thereby reducing the potential for errors and the time required for data entry. Future methodological development plans include transitioning other paper forms such as the teamwork and taskwork questionnaires to computer interfaces.

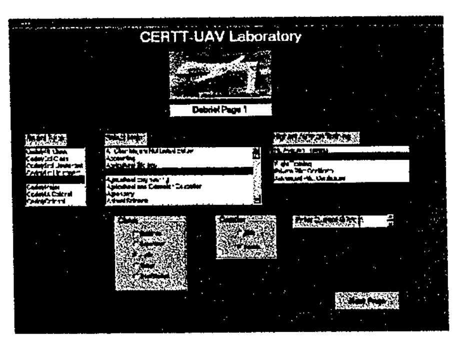


Figure 12. Debriefing questionnaire software

Metrics

Team knowledge metrics. As previously noted, the metrics that have traditionally been used to assess team knowledge are not appropriate for teams of individuals with different knowledge backgrounds (i.e., heterogeneous teams). The traditional metrics include measures of intrateam similarity and overall accuracy. Intrateam similarity assesses the pairwise similarity of the knowledge elicited from individual team members. It has been traditionally assumed that shared mental models imply high intrateam similarity (Cooke, et al., 2000). However, this is probably not the case for heterogeneous teams. Such teams may score low on intrateam similarity simply because they bring different background knowledge to bear on the situation. In addition, intrateam similarity metrics do not take accuracy of the knowledge into account. In some cases, team knowledge has been assessed by aggregating the knowledge accuracy of individual team members, where accuracy is based on comparison to some all-knowing referent (Cooke et al., 2000). Again, the assumption that each team member's knowledge should be assessed by comparison to a single referent is not suited to heterogeneous teams.

Therefore, in addition to using the traditional metrics of intrateam similarity and overall accuracy, we developed some additional metrics that may be better suited to heterogeneous teams. In addition, these metrics can provide more information on the acquisition of knowledge by a team in which different knowledge bases are required. Specifically, we developed team knowledge metrics of positional accuracy and interpositional accuracy. These metrics should reflect mastery of knowledge specific to a particular team position. Thus, in addition to scoring each team member's elicited knowledge against a single all-knowing referent or key (i.e., overall accuracy), that knowledge is also scored against three role-specific keys. In this way, measures of "role" or "positional" accuracy, as well as "interpositional" accuracy (i.e., interpositional knowledge (IPK) or knowledge of roles other than their own) can be determined for each individual. Team accuracy (overall, positional, or interpositional) was the mean accuracy across team members.

Proportion of agreement metric. In our early measurements of team process behavior, we relied on judgments of process made by two or more independent judges. Judgments were made along continuous scale. In order to assess interjudge agreement we developed an index for measuring degree of agreement among scaled variables with a finite range (e.g. human raters assigning values of 1-7). The scaled proportion of agreement index (Po (scale)) is computed by:

$$[\text{Sum}\ (1 - |X1 - X2|/Rg)]/N,$$

where X1 and X2 represent the variables of interest, Rg represents the maximum possible score minus the minimum possible score, and N represents the number of trials (Kiekel, 2000). This statistic expands upon the dichotomous proportion of agreement index (Po) and upon Cohen's kappa, by allowing for an interval scale, while maintaining parsimonious interpretation as a proportion.

The proportion of agreement index also turns out to be a useful way of assessing agreement among team members for scale judgments made. Such team member agreement is relevant to the construct of Shared-Knowledge.

Summary: Methodological Developments

Methodological developments during this effort have ranged from the development of specific quantitative metrics for assessing agreement and team knowledge to hardware and software tools that facilitate recording and analysis of team data collected in the UAV–STE context. Many of these methods and tools have been iteratively refined as they have been tested in pilot studies and the two studies reported here. As we report in the empirical studies that follow, many of the methods have been successful at predicting team performance and distinguishing teams in other ways. Our efforts on methodological development and evaluation are ongoing and future iterations of existing methods and developments of new methods should provide a rich source of information about team cognition and performance.

EMPIRICAL STUDIES

The two empirical studies under this effort were conducted to 1) test the validity of the newly developed team knowledge measures, as well as our versions of the more traditional team performance and process measures (Studies 1 and 2) and 2) using these measures investigate the relation between team cognition and team performance during task acquisition (Study 1) and under conditions conducive or not conducive to knowledge sharing (Study 2). In both studies, the validity of team knowledge measures was assessed in terms of the ability of the measures to predict team performance, process, and situation awareness.

Study 1 was a long-term acquisition study in the context of the team UAV task. Eleven teams of three Air Force ROTC cadets participated in three experimental sessions lasting from three to six hours. During these sessions teams were trained on the task and were observed as they performed ten 40-minute missions. During the missions team performance, team process behavior, and team situation awareness were measured. In addition, long-term team knowledge regarding both taskwork and teamwork were measured apart from the task in four sessions. This study was designed to evaluate a number of different approaches to measuring team knowledge and to examine the development of team performance, process, SA and knowledge as team skill was acquired over the ten missions. The patterns by which team performance, process, situation awareness, and knowledge are acquired may shed light on sequential dependencies among components of team performance and in addition, provide useful information about the point at which asymptotic performance is reached in this synthetic task.

In Study 2, 18 teams of three ROTC cadets participated for a total of five 40-minute missions. Teams were randomly assigned to either the Shared-Knowledge or Nonshared-

Knowledge condition. Teams in the Shared-Knowledge condition were encouraged in a number of ways to share information and learn about the their teammates' positions. Team members in the Nonshared-Knowledge condition were isolated during breaks and not provided with opportunities to look at teammate screens or exchange information about their positions. During the missions team performance, team process behavior and team situation awareness were measured. In addition, long-term team knowledge regarding both taskwork and teamwork were measured apart from the task in three sessions. In this study, in addition to again investigating the validity of the knowledge measures, we were interested in the effect of the shared knowledge manipulation on team performance, process, situation awareness, and knowledge.

In addition to these main objectives, data were also collected in the context of both studies to address several other research questions of secondary importance. These questions include:

1. How do social and demographic factors such as in-group/out-group differences on a team affect team performance (Studies 1 & 2)?
2. How does intragroup trust change as a function of team performance acquisition (Study 1)?
3. How does communication (intragroup flow and content) change with team performance acquisition (Study 1)?
4. How does an ROTC team training seminar impact team performance (Study 1)?
5. How does leadership relate to team performance and cognition (Study 2)?

Analyses directed toward answering these questions are ongoing with the exception of numbers 3 and 4. In regard to the ROTC training, there was no effect of a short training seminar administered to some of the teams during the interval between the second and third session of Study 1. Although there were no hints of a difference, inadequate statistical power may be partially responsible. Communication analyses are ongoing, and the results obtained thus far are presented later in this section.

Study 1: Method

Participants. Eleven three-person teams of Air Force ROTC cadets voluntarily participated in three (3-5 hour) sessions of this study. Individuals were compensated for their participation by payment of $6.00 per person hour to their ROTC organization. In addition, the three team-members on the team of with the highest mean performance score were each awarded a $50.00 bonus.

Equipment and materials. The study took place in New Mexico State University's CERTT (Cognitive Engineering Research on Team Tasks) Lab configured for the UAV team synthetic task described above. For most of the study, each participant was seated at a workstation consisting of two computer monitors (one View Sonic monitor connected to an IBM PC 300PL and one Cyberesearch Industrial Workstation), a Sony video monitor that could present video from a Quasar VCR, a keyboard, a keypad, and a mouse for input. Participants communicated with each other and the experimenters using

David Clark headsets and a custom-built intercom systems designed to log speaker identity and time information. The intercom enabled participants to select one or more listeners by flipping toggle switches.

Two experimenters were seated in a separate adjoining room at an experimenter control station consisting of another IPB PC computer and View Sonic monitor, headsets for communicating with participants, and Panasonic monitors for video feed from ceiling-mounted Toshiba cameras located behind each participant. In addition, a fourth camera captured information from the entire participant room. From the experimenter workstation, the experimenters could start and stop the mission, query participants together or individually, monitor some of the mission-relevant displays, observe team behavior through camera and audio input, and enter time-stamped observations. Video data from cameras was recorded on a Quasar VCR. Audio data from the headsets was recorded on an Alesis digital recorder as well as to the VCR. In addition, custom software recorded communication events in terms of speaker, listener, and the interval in which the push-to-talk button on the microphone was depressed (see methodological developments section for more detail).

Custom software (seven applications connected over a local area net) also ran the synthetic task and collected values of various parameters that were used as input by performance scoring software (see methodological developments section for additional detail). A series of tutorials were designed in Powerpoint for training the three team members (see methodological developments section for additional detail). Custom software was also developed to conduct tests on information in Powerpoint tutorials, to collect individual and consensus taskwork relatedness ratings, and to collect demographic and preference data at the time of debriefing (see methodological developments section for additional detail).

In addition to software, some mission-support materials (rules-at-a-glace for each position, two screen shots per station corresponding to that station's computer displays, and examples of good and bad photos for the PLO) were presented on paper at the appropriate workstations. Other paper materials consisted of the consent forms, debriefing form, a checklist of skills for training, forms for experimenter recording of SA and process, a trust survey, and teamwork and taskwork questionnaires.

Measures. Performance, process, situation awareness and knowledge measures are the focus of this paper. Demographic, preference, trust, video, and communication data were also collected, however, they are secondary to the other measures that are the focus of this report.

Team performance was measured using a composite score based on the result of mission variables including time each individual spent in an alarm state, amount of fuel used, amount of film used, number of targets successfully photographed, and number of critical waypoints visited. Penalty points for each of these components were weighted a priori in accord with importance to the task and subtracted from a maximum score of

1000. Missed targets were weighted four times that of fuel and film used and alarm time and critical waypoints missed were weighted two times fuel and film used.

Team process behavior was scored independently by each of the two experimenters. For each mission the experimenters observed team behavior and responded to a series of nine yes/no questions (see Appendix A) regarding team behaviors that did or did not occur at designated event-triggers in each mission (e.g., DEMPC communicated upcoming target information at the appropriate time to the AVO and PLO). Team process was simply the proportion of the nine process questions that were observed by each experimenter.

Team situation awareness was measured using three SPAM-like (Durso, et al., 1998) queries administered at three randomly selected 5-minute intervals during each mission. One of the experimenters administered the queries to each individual in turn (See Appendix B). Order in which individuals were queried was also random. The three queries asked 1) a prediction regarding the number of targets out of nine successfully photographed by the end of the mission, 2a) the team member or members that they would communicate with next and 2b) the topic of that communication, and 3) the number of targets out of nine successfully photographed thus far. The experimenter also recorded the correct response to these queries once known and this key was used to score the four responses for accuracy. Team accuracy scores were based on the average accuracy of team members. For the second query, this was simply the proportion of correct responses (1 or 0) averaged across the three team members. For the first and third queries, this was the absolute value of the deviation from correct, divided by 9 possible targets and subtracted from 1. Responses to the first and third query were also scored for intrateam similarity. Team similarity was the average of all the pairwise similarities (i.e., converse proportions of absolute deviations) of the three team members. Intrateam similarity was not meaningful for the second query. Thus, there were a total of six situation awareness metrics; two based on similarity and four based on accuracy.

Team knowledge was measured in four separate sessions by four methods: teamwork questionnaire, taskwork questionnaire, taskwork ratings, and taskwork consensus ratings. The teamwork questionnaire (see Appendix C) consisted of a three-part question in which the individual was asked to indicate if directed pairs of team members (e.g., AVO → PLO) pass information in the specified direction. The second part of the question asked them to identify the nature of the information for those communication links identified. The third part asked them to consider any sequential constraints in the timing of the information.

The taskwork questionnaire (see Appendix C) asked individuals to analyze the task starting with the main goal and breaking this up into subgoals and tasks. The next part of this questionnaire asked individuals to associate team roles with each of the tasks and then to indicate any sequential constraints in tasks.

The taskwork ratings consisted of eleven task related terms: altitude, focus, zoom, effective radius, ROZ entry, target, airspeed, shutter speed, fuel, mission time, and photos. All possible pairs of these terms were presented in one direction only, one pair at a time. Pair order was randomized and order within pairs was counterbalanced across participants. Each team member rated the relatedness of each pair on a 1-5 scale with anchors that ranged from slightly related to highly related. There was also an option of unrelated.

Taskwork consensus ratings consisted of the same pairs as taskwork ratings (randomly presented), however the ratings were entered as a team. For each pair, the rating entered in the prior session by each team member was displayed on the computer screen of that team member. The three team members discussed each pair over their headsets until consensus was reached.

Knowledge measures were all scored for accuracy and intrateam similarity. Individual accuracy scores and pairwise measures of response similarity were averaged across team members. For the two rating tasks, data were submitted to KNOT (using parameters *r*=inf. And *q*=n-1) in order to generate Pathfinder networks (Schvaneveldt, 1990). These networks reduce and represent the rating data in a meaningful way in terms of a graph structure with concept nodes standing for terms and links standing for associations between terms. A referent network generated by the experimenters served as the key, and similarity of any one network to this referent in terms of the proportion of shared links was used as a measure of accuracy. In addition, the individual task ratings were scored not only against a key representing overall knowledge, but also against role-specific keys. In this way, measures of "role" or "positional" accuracy, as well as "interpositional" accuracy (i.e., interpositional knowledge (IPK) or knowledge of roles other than their own) could be determined. See Appendix D for overall and positional referent networks. Team accuracy was the mean accuracy across team members. Intrateam similarity was measured using the proportion of shared links for all intrateam pairs of two individual networks (i.e. the mean of the three pairwise similarity values among the three networks).

Procedure. The study consisted of three sessions. Sessions 1 and 2 lasted approximately 5.5 hours each and were separated by a 24-48 hour interval. Session 3 lasted 3.5 hours and followed Session 2 by a lapse of 4 to 8 weeks. During this time seven of the 11 teams participated in a team strategic training seminar offered by the ROTC for the purpose of a separate secondary research question.

In the first session the three participants were randomly assigned to one of the three task positions: AVO, PLO, or DEMPC. Team members retained these positions within the same team for the remainder of the study. The team members were given a brief overview of the study and then were seated at their workstations for training. Team members studied the three Powerpoint training modules at their own pace and were tested with a set of multiple-choice questions at the end of each module. If responses were incorrect, they were instructed to go back to the Powerpoint tutorial and correct their answers. Experimenters provided assistance and explanation if their second response was also incorrect (see methodological developments section for additional detail). Once all team members completed the tutorial and test questions, a mission was started and experimenters had participants practice the task, checking off skills that were mastered (e.g., the AVO needed to change altitude and airspeed, the PLO needed to take a good photo of a target) until all skills were mastered. (See Appendix E for the check list of skills.) Again, the experimenters assisted in cases of difficulty. Training took a total of 1.5 hours.

After a short break the first 40-minute mission began and was completed at the end of the 40-minute interval or when team members believed that the mission goals had been completed. Knowledge measures were then administered in the following order: taskwork ratings, taskwork consensus ratings, taskwork questionnaire, and teamwork questionnaire, and were followed by another short break. Missions 2 and 3 were then administered in the same manner as Mission 1. The second session consisted of Mission 4, followed by knowledge measurement just as before, a break, Missions 5 and 6, a break, Mission 7, and the third knowledge measurement session. Unlike the other missions Mission 7 used a different task scenario. Session 3 started 4-10 weeks following Mission 7, with Missions 8 and 9, followed by a short break, knowledge measurement Session 4, Mission 10, and debriefing. Mission 10 used the same scenario as Mission 7, which differed from the other 8 missions.

Study 1: Results

Overview of analyses. One team did not complete Session 3 and due to equipment malfunctions another team has no performance data recorded for Mission 10. Therefore there are performance, process, and SA data missing for 1 or 2 teams for Missions 8 through 10.

There was adequate agreement between the two experimenters on the team process questions. Agreement between raters was assessed by computing the proportion of agreement between raters across the nine process questions for each team, each mission, and overall. Overall proportion of agreement was .9 (range from .83 to .97). Therefore, ratings were averaged for all cases in which two raters each assigned a score.

A cluster analysis of accuracy and similarity results for the three SA queries was used to identify meaningful groupings of the six metrics. This resulted in four clusters: 1) accuracy to Queries 1 and 3, 2) similarity for Queries 1 and 3, 3) accuracy on the to whom answer to Query 2, and 4) accuracy on the topic answer to Query 2. These were used as indices of team SA.

Finally, due to the use of a small sample of eleven teams, extensive across-team variation, and an objective of identifying any potentially interesting measures or effects at the expense of possible Type I errors, we considered α-levels of $\leq .10$ statistically detectable. Reported correlations of team measures were also based on eleven teams (or fewer for those missions associated with the two teams with missing data) and therefore nine degrees of freedom for which correlations of .52 and higher are required for two-tailed significance at the $p = .10$ level, though we recognize that correlations somewhat lower nonetheless predict a substantial proportion of the variance (Cohen, 1994; Wickens, 1998).

Task acquisition. The team performance score ranged from 353 to 952 with an overall mean of 822 and standard deviation of 74.2. As might be expected and as shown in Figure 13, the standard deviation was greatest for the first and last three missions (range from 106 to 159) and was lowest for the four middle missions (range from 24 to 51). As seen in Figure 14, across the 11 teams, performance improved in general from Mission 1 (M = 510) to Mission 10 (M = 881) (t (8) = 6.70, p < .001), reached asymptote at Mission 4 and then dipped at Mission 8, which was the first mission after the extended break between Sessions 2 and 3. Figure 14 also shows that this drop in performance was greatest for the lowest performing teams. Interestingly, team performance did not suffer as a result of the change in scenario that occurred for Missions 7 and 10 (M = 897 for Missions 7 and 10 and M = 894 for Missions 4, 5, 6, and 9).

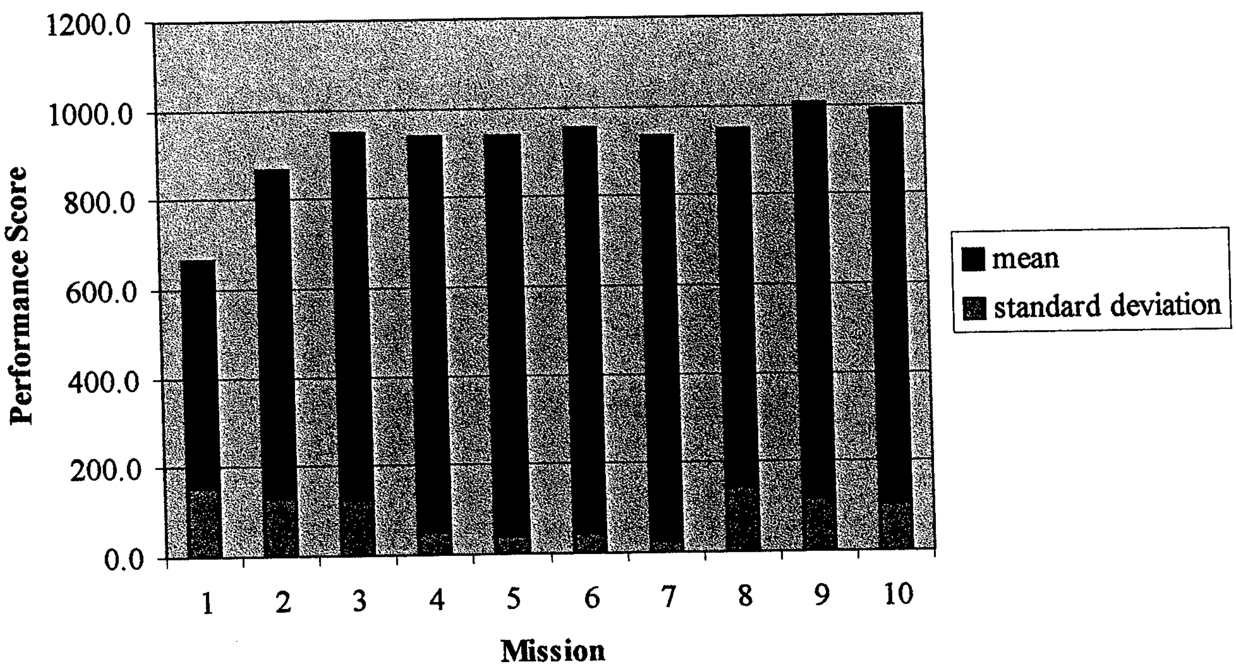


Figure 13. Mean composite performance scores and standard deviations across teams for each mission.

The means for the team process behavior score revealed a pattern of acquisition similar to that for performance, but this was not statistically detectable (F (9,87) = 1.62, p =.122, η^2=.143). The biggest improvement in consecutive missions occurred between Missions 1 and 2 (.74 to .82), but it was also not detectable (t (10) = 1.38, SE = .064, p =.199). However, due to the decrease in error variance over time, a drop of the same magnitude between Missions 8 (M = .87) and Mission 9 (M = .78) was detectable (t (9) = 4.0, SE =.022, p = .003). Mean team performance, process, and SA across the ten missions are shown in Table 2.

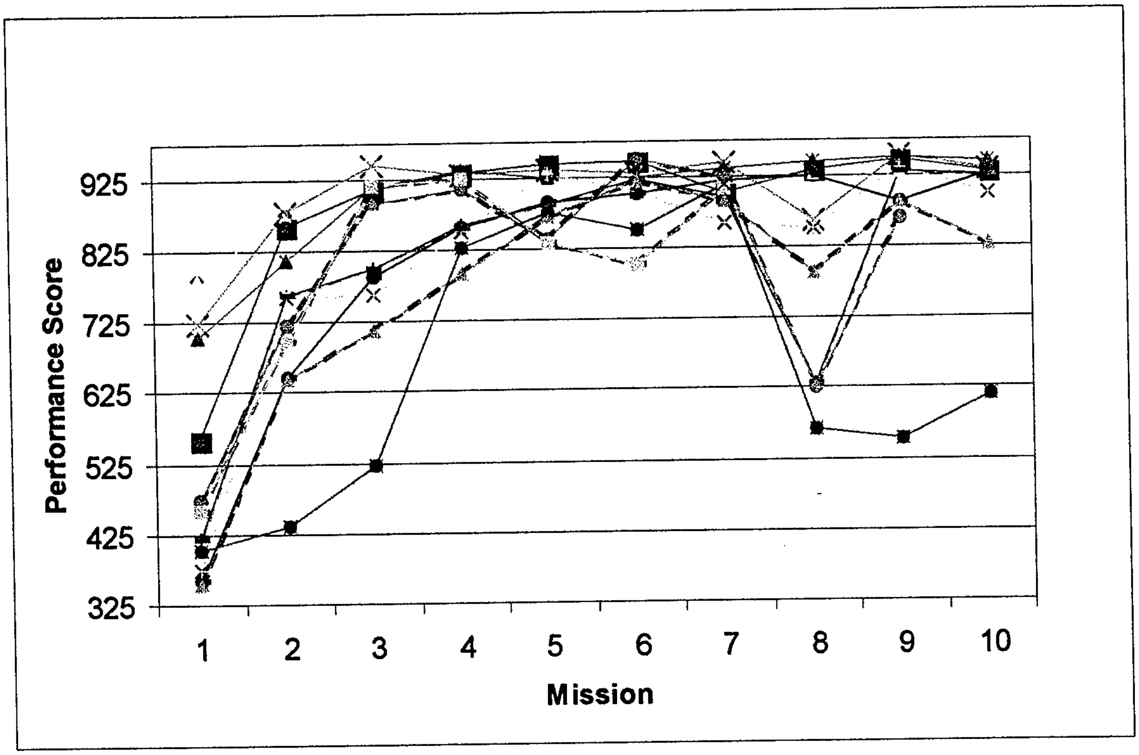


Figure 14. Composite performance scores for each team across each of the 10 missions. Missions 7 and 10 were associated with a scenario different from the other missions.

Table 2. Mean team performance scores, team process scores, and team SA measures across the 10 missions.

MISSION	**Team Performance**	**Team Process**	**Team SA Query 1 & 3 accuracy**	**Team SA Query 1 & 3 similarity**	**Team SA Query 2 - who?**	**Team SA Query 2 - topic?**
Mission 1	509.5	.735	.788	.854	.788	.583
Mission 2	735.3	.823	.868	.889	.792	.558
Mission 3	821.8	.832	.886	.881	.798	.536
Mission 4	885.9	.849	.940	.928	.843	.546
Mission 5	896.6	.859	.956	.970	.783	.458
Mission 6	908.2	.843	.983	.970	.758	.508
Mission 7	910.0	.864	.959	.937	.800	.500
Mission 8	805.2	.867	.887	.942	.704	.509
Mission 9	883.7	.778	.936	.959	.783	.625
Mission 10	881.3	.815	.943	.935	.905	.619

Responses to SA-Query 2 (concerning to whom the individual would talk to next and about what) did not change in any discernable way over time. The other team SA queries, however, did change and in a way that paralleled performance (See Table 2). Accuracy on these queries generally improved from Mission 1 to 10 (.79 to .94

respectively; t (8) = 3.875, p =.005), peaked at Mission 4 (M = .94) and dropped at Mission 8 (M =.89). Also there was no difference between the standard mission scenario and the new one associated with Missions 7 and 10. Intrateam response similarity for SA-Queries 1 and 3 also increased overall (Mission 1 M= .85, Mission 10 M= .94, t (9) = 4.66, p =.001), peaked at Mission 4 (M=.93), and showed no effect of novel scenarios associated with Missions 7 and 10. There was, however, no discernable drop in team SA similarity at Mission 8.

The four knowledge measures showed no significant changes over time with the exception of the teamwork questionnaire, which showed general improvement in team accuracy across the four sessions (M = .53, .66, .71, and .65, respectively; F (3, 29) = 3.083, p = .043, η^2 = .242). Knowledge as measured by this questionnaire seemed to change most drastically between Session 1 and Session 2 (M = .66; t (10) = 2.08, p = .065), which also corresponded to the Mission 4 asymptote seen in the performance data.

How well do measures predict team performance? Team SA (averaged across the 10 missions), as measured by Queries 1 and 3 (accuracy and similarity), correlated reliably with mean team performance (also averaged across the 10 missions) (r(11) = .88, p < .0001 and .72, p = .013, respectively). Multiple regression analysis indicated that most of the predictive power was derived from the Query 1 and 3 accuracy measure (t(10) = 2.91, p = .02). The team process behavior measure did not correlate reliably with performance (r(11) = .132), although several of the individual questions were correlated with performance for the asymptotic Missions 4 through 7.

Critical for the assessment of the validity of knowledge measures is the degree to which they correlate with measures of team performance and to a lesser extent team SA and team process behavior. For correlations between knowledge and performance, data from Knowledge Session 1 were used because 1) with the exception of the teamwork questionnaire there was little difference across sessions, and 2) for some measures across-team variance increased dramatically after Knowledge Session 1 which may indicate that participants took the knowledge task less seriously after the first session. (This is especially true for the taskwork consensus ratings for which the standard deviation of the team accuracy score increased from .04 for Session 1 to .13, .12, and .14 for Sessions 2 to 4 respectively.) Also, given that degree of across-team performance variance changed dramatically across missions (see Figure 13), correlations of knowledge with performance at each mission were computed.

In general, for the various taskwork rating metrics (except role accuracy) and to a lesser extent for taskwork consensus rating accuracy, the measures taken in Knowledge Session 1 were significantly predictive of team performance in the first and last missions (See Table 3). At Knowledge Session 1 greater taskwork rating accuracy, IPK, and intrateam similarity corresponded to higher team performance scores for the early and late missions. Team accuracy and intrateam similarity for both teamwork and taskwork questionnaires generally failed to predict performance.

Table 3. Correlations between knowledge measures taken at Session 1 and performance across the ten missions. Pearson correlations are based on data from eleven teams (df = 9) except for Missions 8 and 9 (10 teams) and Mission 10 (9 teams). With 9 degrees of freedom *r* of .52 is significant at the $p = .10$ level. (* $p < .10$)

MISSION	*Teamwork Questionnaire*		*Taskwork Questionnaire*		*Taskwork Ratings*				**Taskwork Consensus Ratings**
	Accuracy	*Similarity*	Accuracy	*Similarity*	*Accuracy*	*Similarity*	*Role Accuracy*	*IPK Accuracy*	*Accuracy*
Mission 1	-.127	-.174	.143	-.068	.535*	.578*	.186	.232	.377
Mission 2	-.379	.041	-.08	-.380	.839*	.748*	.354	.582*	.252
Mission 3	-.122	.05	-.324	-.532	.769*	.684*	-.048	.605*	.502
Mission 4	.07	-.103	-.321	-.473	.770*	.738*	.004	.613*	.549*
Mission 5	-.382	.084	.279	.001	.485	.505	.443	.368	-.162
Mission 6	-.410	-.017	.548*	.109	.329	.274	.265	.168	-.115
Mission 7	.196	-.053	.232	.105	.085	.094	-.439	-.024	.551*
Mission 8	.037	.419	-.190	-.366	.382	.431	-.116	.555*	.078
Mission 9	-.366	.263	-.215	-.419	.669*	.600*	.146	.557*	.210
Mission 10	-.178	.408	-.271	-.490	.725*	.640*	.045	.677*	.302

Taskwork consensus ratings. This was a new method in which team knowledge was elicited at the team-level. It was assumed that this more holistic approach to measurement would capture not only the collective knowledge of the team members, but also process behaviors of the team that are used in coming to consensus on the ratings (Cooke et al., 2000). Therefore, it was hypothesized that the consensus ratings would be better predictors of team performance than the aggregate taskwork ratings. As indicated in Table 3, accuracy of this measure correlated significantly with team performance at Missions 4 and 7, however it failed to correlate significantly with team process (r (9) = .179, p = .6) and it correlated negatively with team SA: Query 1 and 3 accuracy (r (9) = -.687, p = .02) and similarity (r (9) = -.820, p = .002). Thus, the taskwork consensus ratings, although predictive of performance, did not surpass the aggregate taskwork ratings in their predictive power. However, the accuracy of the taskwork consensus ratings did correspond to the accuracy measure based on individual ratings (r (9) = .522, p = .099), as well as IPK accuracy (r(9) = .659, p = .028).

In order to identify strategies that the teams used to come to consensus in this rating task, the three individual and one team rating for each of the 55 concept pairs were examined for each of the eleven teams. For each pair, the set of four ratings was classified according to one of five rules that mapped individual ratings onto the team rating:

1) all agreed (e.g., AVO=5, PLO = 5, DEMPC = 5, Team = 5)
2) majority (2 out of 3) rules (e.g., AVO = 4, PLO=4, DEMPC = 3, Team =4)
3) leader emerges (e.g., AVO=3, PLO=0, DEMPC=1, Team =3 or AVO=4, PLO=4, DEMPC = 2, Team =2)
4) mid rating (e.g., AVO=0, PLO=3, DEMPC=5, Team =2 or AVO=0, PLO=3, DEMPC=5, Team=3)

5) different from each, and not middle rating (e.g., AVO=5, PLO=2, DEMPC=4, Team=0)

Results of this classification are presented in Table 4. This table illustrates that most teams used strategies 2, 3, and 5 more that 1 and 4. Further, there seems to be little correspondence between the strategies that were used and team performance or process. Experimenters observed that there was very little communication going on during the consensus rating process. Therefore it seems that most teams assumed the strategy to go with majority rule or with the single individual who claimed to have knowledge in the area.

Table 4. Classification of Knowledge Session 1 rating pairs on the basis of mapping individual to team ratings. Highlighted cells occurred more than expected by chance.

Team	Strategy 1	Strategy 2	Strategy 3	Strategy 4	Strategy 5
1	13	3	15	9	15
2	8	16	13	4	14
3	1	21	14	14	5
4	11	8	11	10	15
5	7	16	20	4	8
6	9	15	13	4	14
7	10	10	10	8	17
8	15	16	12	3	9
9	6	12	6	6	25
10	6	12	16	6	15
11	5	19	16	4	11

Study 1: Discussion

In general, the results of this study indicate that teams are able to reach asymptotic levels of team performance on the synthetic UAV team task after 1.5 hours of individual training and four 40-minute missions of teamwork. The fact that asymptotic performance was reached at Mission 4 could be a result of either four trials of practice or the 24-48 hour incubation period that occurred between Sessions 1 and 2, or a combination of both. Also the data indicate that the experience acquired seems to readily transfer to a novel scenario. That is, team performance did not suffer a significant decrement with the presentation of novel scenario for Missions 7 and 10.

On the other hand, team performance did suffer from an extended break of four to ten weeks that occurred between Sessions 2 and 3, as indicated by the drop in team performance (and team situation awareness) at Mission 8. In fact, some of the lowest scoring teams never recovered from this drop. It is also the case that those teams with the four lowest team performance scores at Mission 8, also had relatively long breaks (8-9 weeks) between Sessions 2 and 3. Furthermore, the acquisition of team performance on the synthetic task acquisition was paralleled by changes in Team situation awareness and tended to be preceded by process improvements, suggesting that acquisition of effective team process behavior may be a prerequisite to successful team performance and team situation awareness.

Interestingly, the only noticeable knowledge changes over the four sessions occurred for responses to the teamwork questionnaire on which teams improved across the four sessions and tended to asymptote at Session 2, paralleling the fourth mission. Thus, the team performance and team SA asymptote appear to be paralleled by not only team process improvements, but also by an improved understanding of the teamwork aspects of the task (i.e., knowledge of the team roles and information dependencies).

In hindsight, the first knowledge elicitation session that occurred after training and Mission 1 may have been too late to detect any changes in the two taskwork knowledge measures (questionnaire and rating). Possibly, the most significant growth in knowledge of the team and its tasks occurs during the training session as the team is just learning about the mission and how they will work together. By the time the first mission is complete then, much of the team's broad knowledge is solidified. Alternatively, subtle knowledge structure refinement associated with true expert-level performance may require more experience than teams had in this study. Also, as suggested by increasing variance in some of the knowledge measures (i.e., taskwork consensus ratings), it may be the case that fatigue and boredom contributed to increased noise and lack of reliability in the other knowledge measures, masking any knowledge acquisition that was present.

Although the taskwork relatedness ratings and the taskwork consensus ratings demonstrated little improvement over time, the measures taken in the first knowledge session were predictive of team performance. Those teams with greater knowledge accuracy, IPK, intrateam similarity and consensus accuracy in the beginning tended to have higher scores on early and late missions. This pattern indicates that teams with members who understand the task from the perspective of other positions, and therefore have knowledge similar to one another, are the teams with the highest performance.

In general, the taskwork rating measures seem to be valid indicators of team knowledge, compared to the taskwork and teamwork questionnaires that failed to correlate with team performance. In addition, the more holistic taskwork consensus ratings were also predictive of team performance in some missions. The first session knowledge measures were most predictive of performance and were also associated with lower error variance compared to later knowledge sessions. This pattern again suggests that in general, the rating tasks may be most informative upon first administration.

The taskwork consensus rating task was a new measure developed in attempt to capture the holistic aspects of team knowledge that include not only the aggregate of individual team member knowledge, but also the effects of team process behaviors (see Figure 1). Examination of consensus rating strategies suggests that, quite often, if two team members rated two concepts the same, the third team member conformed to their answer. When none of the members initially agreed on a rating, another popular strategy was for one team member, usually the team member that was considered to have the most knowledge or experience with those concepts, to convince the other team members to change their ratings. Interestingly, the team generally did not just choose to rate the concepts somewhere in the middle of all of their answers (averaging), but instead went with the perceived expertise of one or more of their team members. It is difficult to explain the preponderance of Strategy 5, in which teams entered a rating completely different from, and not a mid point of the three individual ratings.

Although the consensus rating task was predictive of performance and also correlated with the individually-based taskwork ratings, it did not surpass the traditional aggregate measure (i.e., taskwork ratings) in predicting performance (see Table 3). This may call into question the value of the consensus ratings. In many scenarios, it will be more difficult to assemble the team and achieve consensus than to ask team members to rate concepts individually, to be averaged later. A counter to this argument is to note that, although the correlation between the consensus and aggregate approaches was moderately high, ($r(9) = .522, p = .099$), it did not approach colinearity. Hence, these two metrics tap different constructs.

There are several logical but untested explanations for the relative weakness of the taskwork consensus rating method. One possibility relates to the increasing error variance associated with rating tasks in general. The fact that the consensus ratings always followed a set of individual ratings may have exacerbated this problem for the consensus ratings. That is, teams were bored and tired and wanted to quickly finish the task. A second, and related, explanation is inability to concentrate on the consensus rating task, brought on by the pressure for off-task social interaction, coupled with the knowledge that the bonus was tied to mission performance and not the rating task. These two hypotheses are in accord with the preponderance of an apparently random social decision scheme during the consensus ratings. Perhaps teams did not take the task seriously, and simply entered ratings until the three matched.

Overall, the results of this study suggest that the taskwork rating tasks provide valid indicators of team knowledge. In particular, the team knowledge metrics used here are appropriate for teams in which members have different roles. Applying these heterogeneous metrics to the data reveal that highest performing teams have members with more knowledge of the tasks from the perspective of roles other than their own. In other words, knowing multiple roles is better than simply knowing your own. Thus, high performing teams seemed to naturally acquire the kind of knowledge that is consistent with cross training. Measures of team knowledge provide a window to some of the cognitive the factors underlying team acquisition of a complex skill and can thus, be valuable in designing and assessing knowledge-based training programs

Study 2: Overview

As another test of our team knowledge measures we were interested in determining whether the measures could distinguish between teams exposed to different levels of a "Shared-Knowledge" manipulation. Specifically, we manipulated several aspects of the training and test environments to either encourage knowledge sharing -- the mutual exchange of information among team members (i.e., Shared-Knowledge condition)-- or in another condition (Nonshared-Knowledge condition), to discourage such exchange.

To encourage knowledge sharing we provided additional Powerpoint training slides depicting the screens of the other team members and describing their general task. We also encouraged these teams to talk about the task between missions. Teams in the Nonshared-Knowledge condition received additional training, but it consisted of a repetition of previous slides describing their own task. In addition, we covered the computer screens in the Nonshared-Knowledge condition so that they could not be viewed by other team members during the break and we located each team member in separate areas during breaks and asked them not to discuss the task between experimental sessions.

We hypothesized that our manipulation would affect knowledge sharing such that the Shared-Knowledge teams should have greater task knowledge from the perspective of other team members than the Nonshared-Knowledge teams. This knowledge difference should be reflected in our measures. Also, if team knowledge affects team performance, then we predict that again team knowledge should correlate with team performance and the Shared-Knowledge teams should also perform at higher levels of performance than the Nonshared teams. Process and situation awareness measures may also show some affect of the manipulation.

This manipulation and the associated predictions are based on results from and general observations of teams in Study 1. This information indicated that team members were eager to exchange information with their teammates on their jobs, information that they had access to, and systems for which they were responsible. Team members were observed not only discussing these issues with one another, but also showing each other the information on their computer screens between missions. In addition, data on taskwork knowledge and its relation to performance suggested that the highest performing teams also knew more about the task from the perspective of the other positions than did lower scoring teams.

Because we had learned in Study 1 that teams reached asymptotic performance on this task by the fourth mission, we had teams participate in only five missions. The fifth mission was added because we might expect some lengthier acquisition times due to our nonshared manipulation.

Study 2: Method

Participants. Eighteen three-person teams of Air Force ROTC cadets voluntarily participated in two (3-5 hour) sessions of this study. Individuals were compensated for their participation by payment of $6.00 per person hour to their ROTC organization. In addition, the three team-members on the team of with the highest mean standardized performance score were each awarded a $50.00 bonus. Teams were randomly assigned to one of two conditions: Shared-Knowledge or Nonshared-Knowledge.

Equipment and materials. The second study used the same equipment and materials as Study 1 with the following exceptions:

- Slight modifications were made to the training materials to include discussion of a concept (effective radius) that the experimenters considered difficult for participants to master in the first study.

- In addition, training materials were modified to include another module. In the Shared-Knowledge condition, this module presented the screens seen by the other two team members and briefly described the task of these team members. In the Nonshared condition, the module contained a review of several previously-viewed slides focusing only on the information pertinent to that team member.

- Opaque screen covers were made to be attached with Velcro to the computer monitors in the Nonshared condition.

Measures. Performance, process, situation awareness and knowledge measures are the focus of this paper. Demographic, preference, a connation (a management entrepreneurial concept) questionnaire, video, and communication data were also collected, however, they are secondary to the other measures that are the focus of this report.

Team performance was again measured using a composite score based on the result of mission variables including time each individual spent in an alarm state, amount of fuel used, amount of film used, number of targets successfully photographed, and number of critical waypoints visited. In addition, penalties for seconds in a warning state, as well as for violations in route rules (i.e., priority targets not visited first in a ROZ box, ROZ entry and exits not visited) were also included in this composite score. Penalty points for each of these components were weighted a priori in accord with importance to the task and subtracted from a maximum score.

Team process behavior was scored this time by a single experimenter using the checklist approach of the first study. Several modifications were made to these questions and several items were dropped that proved problematic or uninformative (see Appendix A). As before team process was simply the proportion of the six process questions that were observed by the experimenter.

Team situation awareness was measured as in Study 1 using three SPAM-like (Durso, et al., 1998) queries administered at three randomly selected 5-minute intervals during each mission. One of the experimenters administered the queries to each individual in turn (See Appendix B). Order in which individuals were queried was also random. The queries included the first and third queries used in Study 1, dropped the second set of queries that was uninformative, and replaced it with a query about the next target that was scheduled to be photographed. Team situation awareness accuracy and similarity was also scored as in Study 1. Accuracy was the mean proportion of accuracy responses across the three team members and similarity was the mean of the proportion of pairwise response similarities among the three team members.

Team knowledge was measured in three separate sessions by three methods: teamwork questionnaire, taskwork ratings, and taskwork consensus ratings. The teamwork questionnaire (see Appendix C) was modified only slightly from the one used in the first study. The taskwork questionnaire, which was uninformative in Study 1, was not used in Study 2. The taskwork ratings and taskwork consensus ratings were identical to those measures used in Study 1. Further the scoring of accuracy and similarity, as well as the heterogeneous team knowledge metrics was also the same as that of Study 1.

Procedure. The study consisted of two sessions. Session 1 lasted 4.5 hours and was separated by Session 2 by a 24-48 hour interval. Session 2 lasted 3 hours. Session 1 consisted of training, Knowledge Session 1, one 40-minute mission, a 15- minute break, some "pre-test" questionnaires regarding their team competence, another 40-minute mission, and Knowledge Session 2. Session 2 consisted of three 40-minute missions with a 15-minute break between the second and third, Knowledge Session 3 and a set of debriefing questions. Knowledge measures were always administered in the following order: taskwork ratings, taskwork consensus ratings, and teamwork questionnaire. All other procedures were identical to those of Study 1 with the following exceptions:

- Teams were randomly assigned to either the Shared-Knowledge or Nonshared-Knowledge condition. Care was taken to keep team members uninformed of the nature of the manipulation or the labels of the conditions.

- All teams were exposed to an additional training module. The information presented in that module (other team member task information or review of previous information) was dependent on condition.

- Team members in the Shared-Knowledge condition spent break times in the same room and were encouraged by experimenter instruction to discuss the task. Team members in the Nonshared-Knowledge condition were seated in separate areas of the laboratory, were unable to discuss anything with each other during the break, and were prevented from seeing the computer screens by screen covers. These teams were also asked not to discuss the task between sessions.

Study 2: Results

As in Study 1 to the use of a small sample of eighteen teams (eight in the Nonshared condition, ten in the shared), extensive across-team variation, and an objective of identifying any potentially interesting measures or effects at the expense of possible Type I errors, we considered α-levels of $\leq .10$ statistically detectable (Cohen, 1994; Wickens, 1998). All repeated measures tests were analyzed as MANOVA, because some of the tests severely violated the assumption of sphericity. Sphericity-corrected approximate F's are no easier to read, and somewhat less accurate, than multivariate tests. Hence acquisition and condition by acquisition were tested using Hotelling's Trace, the most similar multivariate criterion to the univariate F. In the interest of clarity, trace values are not reported.

Task acquisition. Although this study was not focused on the patterns or characteristics of task acquisition, acquisition data were analyzed in order to select the most meaningful missions or combinations or missions for further analysis. In general, post-asymptotic missions were considered to be the most meaningful and stable indicators of team behavior. In addition, changes in variance across missions or knowledge sessions were also considered in these judgments.

A change across missions was detected for the composite team performance score, ($F(4, 13) = 18.94, p < .001, \eta^2 = .854$). Specifically, an improvement was found between Mission 1 ($M = 367.244$) and Mission 5 ($M = 831.75$), $F(1, 16) = 82.17, p <$ 0.001. The interaction of missions with condition was not statistically detectable for performance score.

We conducted a trial-to-trial test to identify the acquisition asymptote and it indicated that performance continued to improve across trials, but by smaller and smaller increments (see Figure 15), with shared means consistently lying just below Nonshared means. All F's have *df* (1, 16). From Mission 1 to 2, $F = 46.30, p < .001, \eta^2 = .74$; from 2 to 3, $F = 7.75, p = .013, \eta^2 = .33$; from 3 to 4, $F = 5.31, p = .035, \eta^2 = .25$; from 4 to 5, $F = 3.73, p = .071, \eta^2 = .19$. The test between 4 and 5 is statistically detectable, but we suspect that, because of the monotonic drop in effect size, a sixth mission would not have differed from Mission 5. We therefore treat Mission 5 as an asymptote, the same approximate asymptote as for Study 1.

A more rigorous test of the asymptote was also conducted. We modeled performance as a function of overall team performance and the inverse of mission number. Hence, the parameters estimated were the mean of each team across all missions and a weight for the inverse of missions. Our final model was

$$\text{Perf} = -578.51/\text{Mission} + \text{TeamMean} - 1/5 * >_{i=1}^{5}(-578.51/\text{Mission}_i),$$

where Mission is the actual mission number, between 1 and 5 for the observed data. The model correlated adequately to the data, $AdjR^2 = .977$. Estimated extrapolations to Mission 6 did not differ from Mission 5 data, ($F(1, 17) < 1$). Based on these estimates, we can treat Mission 5 as an asymptote. In addition, performance decreases monotonically across trials (ratio of maximum variance to minimum = 6.56).

Finally, we tested the hypothesis suggested by Study 1 that the break between sessions of 24-48 hours caused a major jump in the progression of performance scores due to an incubation effect. In Study 2 the session break occurred between Missions 2 and 3. Figure 15 indicates no major jump between Missions 2 and 3. The η^2 between 2 and 3 is comparable to that between 3 and 4 (Fisher's z' test of comparison = .026) or 1 and 2 (Fisher's z' test of comparison = .168). However, the variance of the Shared-Knowledge condition appears to be more influenced by the break between sessions than the Nonshared group (Shared variance = 35145.30, Nonshared = 5632.65, ratio = 6.24).

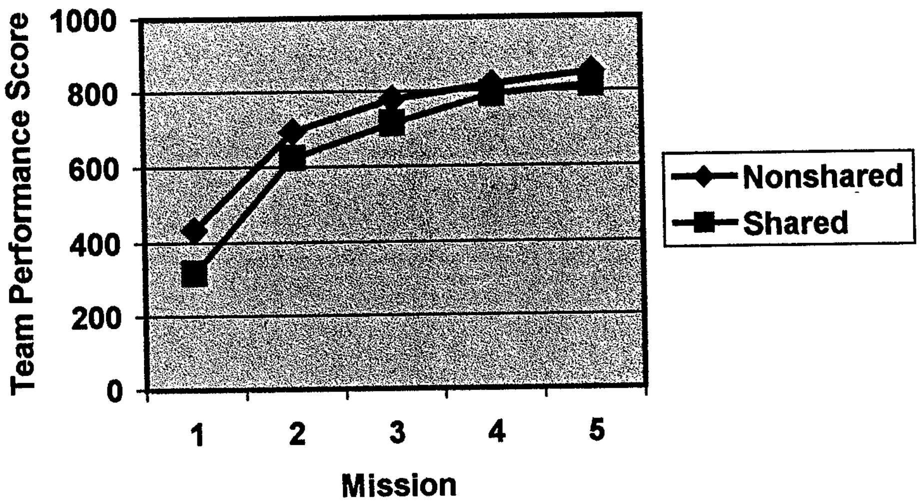


Figure 15. Composite team performance scores for Shared-Knowledge and Nonshared-Knowledge conditions across each of the 5 missions.

No acquisition was detected for team process scores, nor was there an interaction with condition, $F(4, 13) = 1.296, p = .322, \eta^2 = 0.285$. Therefore future analyses of team process averaged process scores across all five missions. Process is continually increasing in variance (ratio of maximum variance to minimum = 3.84), with a high variance at Mission 3, right after the session break

For team SA accuracy, acquisition was detected, $F(4, 13) = 9.34, p = .001, \eta^2 = .74$, but no interaction with condition. Acquisition is an upward incline with no asymptote. The mean for Mission 1 is lower than that for Mission 2, $F(1, 16) = 7.70, p = .014, \eta^2 = .325$. Then there is a plateau between Missions 2 and 4 (both F's < 1). Finally, there is an increase from Mission 4 to 5, $F(1, 16) = 5.25, p = .036, \eta^2 = .247$. This suggests that the best estimate for an asymptote is Mission 5. Further, team SA accuracy shows no change in variance, with the maximum to minimum ratio at 2.13.

For team SA similarity, acquisition was similarly detected, $F(4, 13) = 11.59, p = .001, \eta^2 = .78$, as well as no interaction with condition.. The acquisition shows a pattern of increasing team SA similarity without asymptote. It was not caused by a trial-to-next-trial increase (all p's < .1, all η^2's < .16). Instead, the mean for Mission 1 (M = .67) was lower than that for Mission 3 (M = .80) ($F(1, 16) = 9.63$, $p = 0.007$), Mission 4 (M = .83) ($F(1, 16) = 9.01$, $p = 0.008$), and Mission 5 (M = .92) ($F(1, 16) = 43.47$, $p < .001$). Mean SA accuracy for Mission 2 (M = .749) was lower than that for Mission 5 ($F(1, 16) = 14.17$, $p = 0.002$), and the mean for Mission 3 was lower than for Mission 5 ($F(1, 16) = 5.67$, $p = 0.03$). The lack of convincing asymptote again indicates that the team SA similarity measurement at Mission 5 is the best asymptote estimate.

The teamwork questionnaire accuracy showed no acquisition and no interaction of condition and acquisition, F's < 1. This implies that averaging all sessions is appropriate. Further the maximum variance ratio is 1.95 indicating little change in variance over time. Teamwork questionnaire similarity showed acquisition ($F(2, 15) = 24.59$, $p < .001$, η^2 = .766), but no interaction with condition. Teamwork similarity dropped from Session 1 (M = .38) to Session 2 (M = .36), $F(1, 16) = 10.80$, $p = .005$. It then increased between Sessions 2 and 3 (M = .42), $F(1, 16) = 48.97$, $p < .001$. Because there is change in time, but no asymptote, the most appropriate analysis here is to test each session separately on teamwork similarity. The fact that the maximum variance ratio is 4.15 for teamwork questionnaire similarity (i.e., variance dropping at Session 2) reinforces our intention to analyze the three sessions separately.

For taskwork rating overall and IPK accuracy, taskwork rating similarity, and taskwork consensus rating accuracy there was neither an acquisition, nor an interaction, all F's < 1. However, for taskwork rating role accuracy, there was no significant acquisition effect, but there was an interaction of acquisition and condition, $F(2, 15) = 4.93$, $p = .023$, $\eta^2 = .40$. Mean role accuracy scores at each mission are shown in Table 5 below. As can be seen, the teams in the shared condition had higher role knowledge for all three sessions. There was no condition effect at Session 1 or 3, both p's > .27, $\eta^2 < .1$. The difference was in Session 2, $F(1, 16) = 9.95$, $p = .006$, $\eta^2 = .384$. For all taskwork rating measures, the ratio of maximum variance for a session to minimum was never larger than 2.65. In contrast, for Study 1 the ratio of the lowest variance to the highest for the taskwork knowledge ratings was 12.25. Thus, unlike Study 1, there is not compelling evidence that the teams took only the first session seriously. However, because of the general lack of acquisition effects for the taskwork rating measures, coupled with the odd blip in the role accuracy data at Sessions 2, and the precedent set in Study 1, we will focus the remaining analyses on the taskwork rating measures taken at Session 1 with the exception of role accuracy which we examine at every session.

Table 5. Means for taskwork rating role accuracy across three knowledge sessions for each condition.

	Mean Role Accuracy		
	Session 1	**Session 2**	**Session 3**
Nonshared	0.32	0.29	0.31
Shared	0.35	0.37	0.33
Total	0.34	0.34	0.32

Effect of shared-knowledge manipulation. All measures were tested for condition effects with an ANOVA averaging across all missions/knowledge sessions. Where appropriate, separate ANOVA's were conducted on asymptote estimates resulting from the above analyses of acquisition and variance.

The Shared-Knowledge condition teams scored higher than the Nonshared teams on overall accuracy (Shared $M = .54$, Nonshared $M = .48$, $F(1, 16) = 5.23$, $p = .032$, $\eta^2 = .26$), IPK accuracy (Shared $M = .31$, Nonshared $M = .28$, $F(1, 16) = 4.49$, $p = .05$, $\eta^2 = .219$), role accuracy (Shared $M = .35$, Nonshared $M = .31$, $F(1, 16) = 5.05$, $p = .039$, $\eta^2 = .24$), and intrateam team similarity (Shared $M = .45$, Nonshared $M = .39$, $F(1, 16) = 3.02$, $p = .101$, $\eta^2 = .159$). The role accuracy measure also exhibited the interaction with mission described above indicating that the Shared-Knowledge advantage was detectable only for the second knowledge session.

For all other measures (i.e., team performance scores, team process, team SA accuracy, team SA similarity, taskwork consensus ratings, teamwork questionnaire accuracy, and teamwork questionnaire similarity, no effect was detected for either the average across all missions/sessions, nor at the appropriate asymptote.

How well do measures predict team performance? A table of Pearson correlations is found in Table 6. Only the team SA at Mission 5 correlation with performance at Mission 5, and overall team SA similarity with Missions 2 and 3 are adequate, *rs*= .649, .474, and .422 respectively.

To test for supression effects, we conducted a multiple regression. Condition; team SA accuracy at Mission 5; team process; teamwork questionnaire accuracy; taskwork rating overall, IPK, and role accuracy were entered as predictors. Only team SA was a useful predictor, with *semipartial* $r = .56$, $t(10) = 2.50$, $p = .031$ (all other *semipartial* $|r|$s $< .13$, all p's $> .6$). The model was not adequate with all of the predictors, $F(7, 10) = 1.424$, $p = .296$, *Adjusted* $R^2 = .148$. The same pattern can be seen in the Tables 7 and 8 in which the correlations of the various measures with performance at each of the five missions is broken down by condition (i.e., Nonshared and Shared respectively). Although, team SA is a good predictor of performance for teams in each condition, there are only isolated cases in which knowledge is predictive of performance.

Table 6. Pearson correlations for all measures with performance at all missions. (*p<.10).

		Mission 1	Mission 2	Mission 3	Mission 4	Mission 5
Team SA accuracy	r	0.046	0.179	0.107	0.294	*0.649
at Mission 5	p (2-tail)	0.856	0.477	0.673	0.236	0.004
	N	18	18	18	18	18
Team SA similarity	r	0.216	*0.474	*0.422	0.16	0.25
across all 5 missions	p (2-tail)	0.388	0.047	0.081	0.525	0.316
	N	18	18	18	18	18
Team process across	r	-0.083	0.021	0.057	-0.07	0.001
all 5 missions	p (2-tail)	0.742	0.934	0.822	0.783	0.996
	N	18	18	18	18	18
Taskwork rating IPK	r	-0.197	0.02	-0.03	-0.107	-0.129
at Session 1	p (2-tail)	0.433	0.938	0.907	0.672	0.61
	N	18	18	18	18	18
Taskwork rating role	r	-0.227	-0.001	-0.143	-0.117	-0.027
accuracy at Session 1	p (2-tail)	0.365	0.998	0.571	0.644	0.914
	N	18	18	18	18	18
Taskwork rating role	r	-0.037	-0.01	-0.27	-0.22	-0.118
accuracy at Session 2	p (2-tail)	0.884	0.968	0.279	0.381	0.642
	N	18	18	18	18	18
Taskwork rating role	r	-0.19	-0.056	-0.07	0.018	0.101
accuracy at Session 3	p (2-tail)	0.449	0.826	0.784	0.944	0.689
	N	18	18	18	18	18
Taskwork rating overall	r	-0.225	0.04	-0.157	-0.026	-0.047
accuracy across at	p (2-tail)	0.369	0.874	0.533	0.917	0.853
Session 1	N	18	18	18	18	18
Taskwork rating similarity	r	0.006	.162	.044	.091	.010
at Session 1	p (2-tail)	0.980	0.521	0.862	0.719	0.970
	N	18	18	18	18	18
Taskwork consensus	r	0.158	-0.063	-0.298	-0.223	-0.338
rating at Session 1	p (2-tail)	0.532	0.803	0.229	0.373	0.17
	N	18	18	18	18	18
Teamwork questionnaire	r	-0.083	0.188	-0.172	-0.135	0.176
similarity at Session 1	p (2-tail)	0.744	0.455	0.496	0.592	0.484
	N	18	18	18	18	18
Teamwork questionnaire	r	0.228	0.243	-0.04	-0.236	0.116
similarity at Session 2	p (2-tail)	0.363	0.332	0.876	0.347	0.647
	N	18	18	18	18	18
Teamwork questionnaire	r	-0.014	0.132	-0.173	-0.098	0.257
similarity at Session 3	p (2-tail)	0.956	0.601	0.494	0.698	0.303
	N	18	18	18	18	18
Teamwork questionnaire	r	0.091	0.141	-0.173	-0.027	0.07
accuracy across	p (2-tail)	0.721	0.577	0.494	0.916	0.783
all 3 sessions	N	18	18	18	18	18

Table 7. Pearson correlations for all measures with performance at all missions for Nonshared-Knowledge condition. (*p<.10).

		Mission 1	Mission 2	Mission 3	Mission 4	Mission 5
Team SA accuracy at Mission 5	r	*0.66	*0.749	*0.849	*0.858	*0.752
	p (2-tail)	0.075	0.032	0.008	0.006	0.032
	N	8	8	8	8	8
Team SA similarity across all 5 missions	r	0.323	0.282	-0.151	-0.321	0.022
	p (2-tail)	0.436	0.499	0.72	0.439	0.959
	N	8	8	8	8	8
Team process across all 5 missions	r	-0.072	0.175	*0.668	*0.668	0.365
	p (2-tail)	0.865	0.678	0.07	0.07	0.373
	N	8	8	8	8	8
Taskwork rating IPK at Session 1	r	-0.091	0.094	-0.127	-0.616	*-0.626
	p (2-tail)	0.831	0.825	0.764	0.104	0.097
	N	8	8	8	8	8
Taskwork rating role accuracy at Session 1	r	-0.06	0.189	0.025	-0.441	-0.43
	p (2-tail)	0.888	0.655	0.952	0.274	0.288
	N	8	8	8	8	8
Taskwork rating role accuracy at Session 2	r	0.374	0.545	0.059	-0.03	0.116
	p (2-tail)	0.361	0.163	0.89	0.944	0.784
	N	8	8	8	8	8
Taskwork rating role accuracy at Session 3	r	-0.047	0.086	-0.18	-0.175	-0.064
	p (2-tail)	0.911	0.839	0.67	0.679	0.88
	N	8	8	8	8	8
Taskwork rating overall accuracy at Session 1	r	-0.014	0.194	0.098	-0.416	-0.52
	p (2-tail)	0.974	0.645	0.818	0.306	0.187
	N	8	8	8	8	8
Taskwork rating similarity at Session 1	r	0.148	.046	.038	-.557	*-.766
	p (2-tail)	0.726	0.914	0.929	0.152	0.027
	N	8	8	8	8	8
Taskwork consensus Rating at Session 1	r	0.207	-0.015	-0.39	-0.206	-0.102
	p (2-tail)	0.623	0.972	0.34	0.624	0.811
	N	8	8	8	8	8
Teamwork questionnaire similarity at Session 1	r	-0.068	0.444	-0.234	0.061	0.52
	p (2-tail)	0.873	0.27	0.576	0.887	0.187
	N	8	8	8	8	8
Teamwork questionnaire similarity at Session 2	r	0.336	0.543	-0.116	0.094	0.509
	p (2-tail)	0.415	0.164	0.784	0.825	0.197
	N	8	8	8	8	8
Teamwork questionnaire similarity at Session 3	r	-0.091	0.297	-0.392	-0.048	0.479
	p (2-tail)	0.83	0.475	0.337	0.909	0.23
	N	8	8	8	8	8
Teamwork questionnaire accuracy across all 3 sessions	r	0.253	*0.67	0.272	0.506	*0.731
	p (2-tail)	0.545	0.069	0.514	0.2	0.039
	N	8	8	8	8	8

Table 8. Pearson correlations for all measures with performance at all missions for Shared-Knowledge condition. (*p<.10).

		Mission 1	Mission 2	Mission 3	Mission 4	Mission 5
Team SA accuracy	r	-0.312	-0.023	-0.027	0.143	*0.635
at Mission 5	p (2-tail)	0.38	0.949	0.942	0.694	0.048
	N	10	10	10	10	10
Team SA similarity	r	0.273	*0.572	0.5	0.249	0.334
across all 5 missions	p (2-tail)	0.446	0.084	0.141	0.488	0.345
	N	10	10	10	10	10
Team process across	r	-0.116	-0.089	-0.139	-0.456	-0.246
all 5 missions	p (2-tail)	0.749	0.806	0.703	0.185	0.493
	N	10	10	10	10	10
Taskwork rating IPK	r	-0.269	0.011	0.078	0.313	0.429
at Session 1	p (2-tail)	0.453	0.977	0.829	0.379	0.217
	N	10	10	10	10	10
Taskwork rating role	r	-0.36	-0.109	-0.188	0.206	*0.558
accuracy at Session 1	p (2-tail)	0.307	0.764	0.602	0.567	0.094
	N	10	10	10	10	10
Taskwork rating role	r	-0.2	-0.375	-0.393	-0.351	-0.08
accuracy at Session 2	p (2-tail)	0.58	0.285	0.261	0.32	0.825
	N	10	10	10	10	10
Taskwork rating role	r	-0.234	-0.093	0.037	0.186	0.333
accuracy at Session 3	p (2-tail)	0.515	0.799	0.919	0.608	0.347
	N	10	10	10	10	10
Taskwork rating overall	r	-0.375	0.004	-0.216	0.31	0.493
accuracy at Session 1	p (2-tail)	0.286	0.991	0.549	0.383	0.148
	N	10	10	10	10	10
Taskwork rating similarity	r	-.232	.285	0.052	*.575	*.730
at Session 1	p (2-tail)	0.518	0.425	0.886	0.082	0.016
	N	10	10	10	10	10
Taskwork consensus	r	0.165	-0.083	-0.308	-0.243	-0.537
Rating at Session 1	p (2-tail)	0.648	0.82	0.386	0.498	0.109
	N	10	10	10	10	10
Teamwork questionnaire	r	-0.29	-0.269	-0.382	*-0.594	-0.446
similarity at Session 1	p (2-tail)	0.417	0.452	0.276	0.07	0.196
	N	10	10	10	10	10
Teamwork questionnaire	r	0.163	0.076	-0.017	-0.385	-0.096
similarity at Session 2	p (2-tail)	0.652	0.834	0.962	0.273	0.792
	N	10	10	10	10	10
Teamwork questionnaire	r	-0.26	-0.508	*-0.633	*-0.674	-0.392
similarity at Session 3	p (2-tail)	0.469	0.134	0.05	0.032	0.263
	N	10	10	10	10	10
Teamwork questionnaire	r	0.005	-0.142	-0.287	-0.244	-0.269
accuracy across	p (2-tail)	0.99	0.696	0.421	0.496	0.452
all 3 sessions	N	10	10	10	10	10

Study 2: Discussion

The purpose of the second study was to evaluate the ability of our knowledge measures to reflect knowledge differences introduced through manipulation of the training and task environments. Specifically, teams in the Shared-Knowledge group were encouraged to exchange task information with other team members and were presented with additional information about team member jobs. This could be considered an abbreviated form of cross training (Cannon-Bowers, Salas, Blickensderfer, & Bowers, 1998). Teams in the Nonshared condition on the other hand, were discouraged from this type of information exchange and were not presented with the information about the task from the perspective of the other team members.

Results indicate that the taskwork rating-based measures, with the exception of the taskwork consensus ratings, were able to distinguish the team knowledge in the two conditions. Shared-knowledge teams had more taskwork knowledge than Nonshared teams and specifically, had knowledge with greater degrees of overall, IPK, and role (Session 2 only) accuracy and greater intrateam knowledge similarity than the Nonshared teams. This is not surprising, as the manipulation was intended to affect Shared-Knowledge such that teams in the shared-knowledge condition had more knowledge of the task from the perspective of all of the team members. Having high overall and IPK accuracy and being similar to one another in terms of task knowledge structures can be interpreted as understanding the task from the same point of view – a point of view that included the perspective of each of the three team positions. We call this a "global view" of the task.

Thus, the ability of the taskwork rating measures to differentiate the two conditions is positive support for the validity of these measures. On the other hand, it is puzzling that the Shared-Knowledge manipulation had no effect on performance and that the knowledge measures are not predictive of performance in this study. This pattern of results suggests that our manipulation had intended effects on knowledge, but did not influence performance and likewise, Shared-Knowledge acquired in this study was not related to higher levels of performance.

One possibility is that a global view of the task is only indirectly related to performance. In fact, this indirect relation is consistent with the framework presented in Figure 1, in which collective team knowledge (e.g., taskwork rating accuracy) combines with team process to generate holistic team knowledge that <u>is</u> directly related to team performance. Thus, a good measure of holistic knowledge (apparently not our taskwork consensus ratings) might show manipulation effects in the team holistic knowledge, as well correlation with the team performance. It is not clear what kind of process behavior (or behaviors) is associated with the highest performing teams in our studies, as our measures of process have not been sensitive enough to demonstrate much acquisition effects or correlations with performance. However, it is possible that good teams have both global task knowledge and some undetermined process behavior that together lead to or are simply associated with high performance.

Our manipulation, resulted in an increased global view of the task for the Shared-Knowledge teams, but evidently this was insufficient for high performance. In fact, the manipulation may have interfered with the development of this other component that is needed. Alternatively, direct training of this global view may have bypassed some alternative natural knowledge acquisition process that also leads to the development of the undetermined process component. In general, the manipulation succeeded in teaching the shared-knowledge teams a global view, but interfered with typical task acquisition, thereby compromising team performance.

It may also be relevant that in neither Study 1 nor Study 2 nor other similar team studies (e.g., Cooke, Cannon-Bowers, Kiekel, Rivera, Stout, & Salas, 2000), do the rating-based measures demonstrate development over knowledge sessions, yet they are (except for the current study) predictive of team performance. As mentioned before, this may be due to lack of reliability of the technique over repeated trials. It also could be that the raying-based measures capture more of an aptitude for quickly grasping taskwork knowledge. High performing teams may have aptitudes for quickly understanding the task from the global view. Thus, training in the global view when there is no aptitude, may be successful on the surface (as in "teaching to the test"), but misses the mark on some deeper understanding that is acquired by teams with aptitude. Without the deeper understanding there is no consequence for performance. Based on this explanation, the manipulation succeeded in training one condition to do well on the taskwork rating test, but the training was not at the level that would affect performance.

Thus, in this study we have not only gained support for the validity of our knowledge measures, but have gained a better understanding of the complexity of the knowledge-performance relationship associated with this team task. There are some other findings that are also worthy of note. In particular, the query-based measures of the teams' situation awareness (i.e., the situation model) have been quite successful in both studies in terms of demonstrating acquisition with task experience and correlating with team performance. The fact that these measures are taken on-line with little apparent task interference is also notable.

Communication Analysis

Communications data collected over the headsets during the first seven of the ten missions of Study 1 were analyzed to investigate potential relationships between communication pattern and team effectiveness (measured in terms of composite team performance score across each of the seven missions).

Toward this effort, our consultant, Dr. Clint Bowers along with Florian Jentsch from the University of Central Florida, transcribed the audio from the video tapes and coded the resulting transcripts in terms of communication category (e.g., planning, response, acknowledgement, factual). As indicated in their summary below, their analyses investigated overall communication frequency differences across teams and the seven missions. In addition, they looked at frequencies within the coded categories, as well as sequential patterns consisting of two, three or four statements. As is revealed in their summary of this work below, there were no significant frequencies or patterns distinguishing teams or missions. Correlations that we later ran between overall communication frequency and team performance score across the 11 teams revealed a positive, but nonsignificant correlation of .313. Further, the five teams identified by Bowers and Jentsch as having fewer nontask statements than typical included two of the four lowest scoring teams and the two highest scoring teams.

Executive summary of Bowers and Jentsch. The present study attempted to assess the role of sequential communication patterns in team development. Communications among team members in a synthetic task that required high levels of team interaction for successful performance were investigated. In addition to the analysis of single-statement frequency data, pattern analyses were used to compare two-, three-, and four-statement communication sequences among the teams. Additionally, the occurrence of those communication sequences that use feedback loops was studied.

The data suggested that the teams were surprisingly homogeneous in their communications, both within and across sessions, as well as across teams. Two-statement sequences indicating closed feedback loops were ubiquitous, indicating consistent communications in the teams.

The homogeneity of the communications, when taken in combination with the relatively structured nature of the task across sessions and teams, suggested that there were relatively few changes in the teams and differences across teams that manifested themselves in communications. However, a validation of the communication measures as an indication of differences in team processes or performance was achieved. Further analyses should compare whether there are any differences in communication frequencies and sequences related to differences in other process and/or performance measures.

The absence of large variability in the data also promises that relevant experimental manipulations, when introduced in subsequent studies, have a relatively strong chance of showing up in team communications. Towards this end, we recommend that great care be taken in future studies to record and analyze team communications across teams and in all sessions.

Future communications work. These communications analyses, although laborious in nature (i.e., transcribing and coding time) only scratch the surface in terms of capturing the richness of team communications. That is, there is no indication in these analyses of speaker or listener identity or communication flow patterns among team members. This situation, along with some of the methodological developments in the CERTT Lab regarding communication logging, has prompted a new line of research. This new research program (with Nancy Cooke, Peter Foltz, Steven Shope, Preston Kiekel), currently under Office of Naval Research sponsorship, has as its goal the automation of team communication analysis techniques.

This new research capitalizes on the comlog data captured in the CERTT Lab that automatically records speaker, listener, and duration of communication at specified intervals. Thus, the laborious transcription of this information can be bypassed and communication flow or traffic can be studied among team members. In addition to investigating the frequencies and sequences of communication categories, speaker and listener can also be taken into account. Some preliminary analyses using a statistical technique developed in-house for iterative clustering and testing of communications patterns, revealed that high performing teams are associated with more consistent communications traffic than low performing teams. This result, albeit preliminary, suggests that the most meaningful communications data, may require a deeper look than overall frequencies and sequences devoid of speaker identity.

Another facet of our new research program explores the use of Latent Semantic Analyses (Landauer, et al., 1998) for automated coding of communications data once transcribed. We hope that we can apply this technique to communications data to generate coded categories comparable to human coding results. Ultimately, we anticipate that the combination of automated recording of communications flow and Latent Semantic Analysis of content will result in marked efficiency with which communications data can be recorded and analyzed, while at the same time taking advantage of the information richness that exists.

Summary: Empirical Studies

The two empirical studies, as well as several pilot studies, and empirical efforts indirectly related to this effort have served to identify promising methods of measuring team cognition. In general, the taskwork relatedness rating measures taken at the individual level seem to provide useful information about the team's knowledge of the task from the perspective of each team role. The teamwork questionnaires used in these studies reflected knowledge that changed with mission experience, but was not generally associated with team performance. The measure of team situation awareness, on the other hand, seemed to capture the momentary knowledge of the team regarding the mission in progress. This measure was predictive of performance in both studies and unlike the taskwork rating-based measures, was administered during mission performance in the form of experimenter queries randomly interspersed through the mission. Other methods tested in these studies have not been as successful at measuring that which they were intended to measure, including the taskwork consensus ratings, the taskwork questionnaire, and the team process measure. These methods require further iteration or possibly re-conceptualization.

The empirical work has also shed light on the nature of team performance and cognition as situated in the UAV-STE. Teams of ROTC cadets were able to quickly (1.5 hours) acquire the skill that they needed to perform their individual roles and within four 40 –minute missions after training had reached asymptotic levels of performance. Team situation awareness followed a parallel developmental path. Further the teams' skills were not specific to the UAV scenario in so much as performance was unaffected by a novel scenario. Performance was, however, affected by a long break of several weeks, especially for weaker teams. Finally, although the best teams in Study 1 had knowledge that resembled a global view of the task (i.e., from the perspective of all three team positions), attempts to directly train this form of knowledge succeeded in terms of team knowledge acquisition, but had no impact on performance. Thus, it seems that the possession of a global view of the task is only partly responsible for high levels of team performance. It is likely that team process behaviors play a role and that mastery of this component of skill was thwarted by the manipulation in Study 2. In general, the UAV-STE provides a complex and dynamic task environment in which teams can reach proficiency in a reasonable amount of time, yet teams can also be differentiated from each other in terms of their level of skill and concomitant knowledge and process.

CONCLUSIONS

During this three-year effort we have made significant progress toward our long-range objectives. We summarize the objectives and the accomplishments made relevant to each below.

- **Identify needs and issues in the measurement of team cognition.**
  - Reviewed relevant literature on team cognition
  - Developed a framework for understanding team cognition
  - Identified measurement needs in the context of this framework
  - Reported this work in several papers and presentations

- **Develop a military synthetic task environment that emphasizes team cognition.**
  - Continued ongoing development of CERTT Laboratory hardware and software
  - Collected information on the Predator UAV operations task
  - Identified design constraints
  - Designed the UAV-STE
  - Demonstrated the UAV-STE in numerous forums
  - Tested the UAV-STE in empirical studies

- **Develop new methods suited to the measurement of team cognition**.
  - Developed methods and tools to facilitate experimenter control, parameter-setting, manipulation, and observations in the CERTT Lab.
  - Developed methods and tools for the measurement of team performance and cognition.
  - Developed metrics appropriate for assessing team knowledge of heterogeneous teams.
  - Based on some preliminary promising analyses, initiated a program of research to develop automated methods for analyzing team communications data.

- **Evaluate newly developed methods.**

- Evaluated measures of team knowledge, as well as team performance, process, and situation awareness in the context of studies using ROTC cadets in the UAV-STE.
- Iteratively designed and tested methods using results of pilot studies, related studies, and Study 1.
- Identified promising methods to include: taskwork relatedness ratings to assess team knowledge of taskwork, teamwork questionnaire to assess team knowledge of teamwork, situation awareness mission probes to assess team situation awareness, and a composite score to assess team performance
- Identified measures in need of additional iteration or re-conceptualization to include: team process behavior measure, taskwork questionnaire of team knowledge, and taskwork consensus ratings of holistic team knowledge.

- **Apply methods to better understand team cognition.**
    - Applied methods in the context of empirical studies to understand the acquisition of team performance and cognition with UAV-STE mission experience
    - As secondary research questions investigated the effects of trust, social and demographic factors, and leadership on team cognition and performance.

- **Apply methods to evaluate interventions relevant to team cognition.**
    - Applied methods in the context of empirical studies to evaluate a training intervention designed to support Shared-Knowledge.
    - As a secondary research question investigated the effect of an ROTC team training seminar on team cognition and performance.

Details of these various accomplishments can be found elsewhere in this report.

Of critical importance are the implications of this work to the to the mission of the 21st Century US Air Force. Along these lines, we have identified five major implications.

1) The UAV-STE provides a test-bed within which research can be carried out with results directly applicable to Air Force UAV operations and more importantly, to other Air Force command and control tasks and to team tasks that require team knowledge sharing, planning, decision making, and communication. All of these tasks are becoming increasingly important to the mission of the 21st century Air Force and military in general. Because the UAV-STE was developed with the Predator UAV operations in mind, it is assumed that results scale to this task, but in-line with the abstraction process associated with STE development, results obtained in the UAV-STE should scale more broadly to tasks that have similar characteristics. Of course, the ultimate test of

generalizeability rests with field studies in which measures are applied and empirical results are replicated in the context in question. Accordingly, we have kept the field-ability of our results in mind during the course of this research to enable their eventual in-situ testing.

2) The UAV-STE and the CERTT Lab in which it is embedded, are flexible and capable of extension to additional UAV scenarios, to similar command-and-control tasks, to different Air Force team tasks, and even to nonmilitary tasks requiring teams such as Air Traffic Control and operating room scenarios. These kinds of extensions can further enhance the generalizeability of the results and allow explorations of team cognition across very different contexts. Importantly, these modifications can be made within the existing measurement and experimental control infrastructure of the CERTT Lab so that previous efforts in these areas can benefit the new tasks. For example, the communication logger and the analytic methods applied to those data can be readily ported to a different STE. Additionally, these particular methods are portable to field experiments.

3) The CERTT Lab and UAV-STE are also capable of connecting via high-speed internet connections to other Air Force and military STEs and simulations. This situation enables the exploration of team cognition in hierarchical "teams of teams." Further, the teams of teams and even the individuals in the UAV-STE can interact in geographically dispersed settings, facilitating the exploration of the effect of DMEs (distributed mission environments) on team cognition. Indeed, developments toward hierarchical and distributed teams in the UAV-STE context is one area targeted for future work. Again, distributed and hierarchical teams will play an increasingly important role in future Air Force command and control operations.

4) The measures developed under this effort are also designed not only with automation and portability in mind, but also with the nature of typical Air Force teams in mind. That is, Air Force teams like UAV control teams, are heterogeneous with team members bringing to the task diverse backgrounds, knowledge, and skills. Previous measures of team knowledge have focused on team member similarity. This focus may be inappropriate for Air Force and other heterogeneous teams.

5) Under this effort, measures of team knowledge and team situation awareness have been developed that are predictive of team performance. The ability to assess team cognition and predict team performance has far reaching implications for evaluating progress in Air Force training programs and diagnosing the instruction needs during training. Further, understanding the cognition underlying team performance has implications for the design of technological aids to team performance not only in training but, most importantly, in actual military and battle space operations.

There are a number of future directions that we have targeted for this research program, keeping in mind our long-term objectives. In the immediate future we are pursuing a line of research in the UAV-STE context that moves the task into a distributed mission environment. Of particular interest is the effect of such a DME, as is common in

21st Century battlefield, on team cognition and performance. Understanding these effects is paramount to identifying and understanding team performance difficulties in DMEs so that corrective action can be taken. Other factors that we plan to investigate in these or future studies in the UAV-STE include the effects of workload, intrateam familiarity, and turnover on team cognition. Future empirical studies will also be directed toward the testing of our team cognition framework presented in Figure 1.

We are also in the process (with Veridian and Sam Schifflet at Brooks AFB) of testing a connection between CERTT's UAV-STE and an AWACS STE at Brooks. This work will move the UAV-STE into a broader hierarchical team arena and enable investigations of teams of teams and team cognition in this context.

There are a number of directions that we are pursuing in regard to methodological developments. The communication analysis work is ongoing and has several promising directions. The application of multivariate methods to the analysis of the comlog data to represent communication traffic patterns among team members is one such direction. In addition the use of Latent Semantic Analysis to automate the process of content coding of communication data is also promising. Additional methodological work is ongoing to develop a measure of holistic team knowledge, perhaps refining the taskwork consensus rating method used in this effort. We also seek an improved measure of team process behavior. Although team process is not the focus of our work, we believe it plays a significant role in team cognition and thus, its sound measurement is critical.

REFERENCES

Braun, C. C., Bowers, C. A., Holmes, B. E., & Salas, E. (1993). Impact of task difficulty on the acquisition of aircrew coordination skills. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting,* 1262-1266.

Brehmer, B. & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior, 9,* 171-184.

Cannon-Bowers, J. A., Burns, J. J., Salas, E., & Pruitt, J. S. (1998). Advanced technology in scenario-based training. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making Decisions Under Stress,* (pp. 365-374). Washington, DC: American Psychological Association.

Cannon-Bowers, J. A., and Salas, E. (1997). Teamwork competencies: The interaction of team member knowledge skills and attitudes. In O. F. O'Neil (Ed.), *Workforce readiness: Competencies and assessment* (pp. 151-174). Hillsdale, NJ: Erlbaum.

Cannon-Bowers, J. A., Salas, E., Blickensderfer, E., & Bowers, C. A. (1998). The impact of cross-training and workload on team functioning: A replication and extension of initial findings. *Human Factors, 40,* 92-101.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In J. Castellan Jr. (Ed.), Current *issues in individual and group decision making* (pp. 221-246). Hillsdale, NJ: Erlbaum.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Cognitive skills and their acquisition* (pp. 141-189). Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The earth is round (*p*<. 05). *American Psychologist, 49,* 997-1003.

Cooke, N. J., Cannon-Bowers, J. A., Kiekel, P. A., Rivera, K., Stout, R., and Salas, E. (2000). Improving team's interpositional knowledge through cross training. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* and Paper presented at 44th annual meeting of the Human Factors and Ergonomics Society and International Ergonomics Association, July 30-August 4, San Diego, CA.

Cooke, N. J., Kiekel, P.A., Salas, E., Stout, R., Bowers, C, & Cannon-Bowers, J.(in preparation). Measuring team knowledge: A window to the cognitive underpinnings of team performance differences. To be submitted to *Group Dynamics.*

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. (2000). Measuring team knowledge. *Human Factors, 42,* 151-173.

Cooke, N. J. & Schvaneveldt, R. W. (1988). Effects of computer programming experience on network representations of abstract programming concepts. *International Journal of Man-Machine Studies, 29,* 407-427.

Cooke, N. J. & Shope, S. M. (1998). Facility for Cognitive Engineering Research on Team Tasks. Report for Grant No. F49620-97-1-0149, submitted to AFOSR, Bolling AFB, Washington, DC.

Cooke, N. J., Stout, R., & Salas, E. (1997) Expanding the measurement of situation awareness through cognitive engineering methods, *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting,* 215-219.

Davis, J. H. (1973). Group decision and social interaction: A theory of social decision schemes. *Psychological Review, 80(2)*: 97-125.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., & Nikolic, D. & Manning, C. A. (1998). Situation awareness as a predictor of performance in en route air traffic controllers. *Air Traffic Control Quarterly.*

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors, 37,* 65-84.

Foushee, H. C. (1984). Dyads and triads at 35,000 feet. *American Psychologist, 39,* 885-893.

Fowlkes, J., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *The International Journal of Aviation Psychology, 8,* 209-221.

Glaser, R. & Chi, M. T. H. (1988). Overview. In M.T.H. Chi, R. Glaser, and M.J. Farr (Eds.), The *Nature of Expertise* (xv-xxviii). Hillsdale, NJ: Erlbaum.

Gugerty, L., DeBoom, D., Walker, R., & Burns, J. (1999). Developing a simulated uninhabited aerial vehicle (UAV) task based on cognitive task analysis: Task analysis results and preliminary simulator data. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 86-90). Santa Monica, CA: Human Factors and Ergonomics Society.

Gugerty, L., Hall, E., & Tirre, W. (1998). Designing a simulation environment for uninhabited aerial vehicle (UAV) operations based on cognitive task analysis. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (p. 1609). Santa Monica, CA: Human Factors and Ergonomics Society.

Hinsz, V. B. (1999). Group decision making with responses of a quantitative nature: The theory of social decision schemes for quantities. *Organizational Behavior and Human Decision Processes, 80(1)*: 28-49.

Howell, W. C., & Cooke, N. J. (1989). Training the human information processor: A look at cognitive models. In I. Goldstein (Ed.), *Training and Development in Work Organizations: Frontier Series of Industrial and Organizational Psychology, Volume 3,* New York: Jossey Bass, 121-182.

Kiekel, P. A. (2000). Communication of emotion in remote collaborative writing with rich versus lean groupware environments. Unpublished M.A. thesis. New Mexico State University, Las Cruces, NM.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20,* 403-437.

Landauer, T. K, Foltz, P. W. & Laham, D. (1998). An introduction to Latent Semantic Analysis. *Discourse Processes, 25,* 259-284.

Leedom, D. K., & Simon, R. (1995). Improving team coordination: A case for behavior-based training. *Military Psychology, 7,* 109-122.

Martin, E., Lyon, D. R., & Schreiber, B. T. (1998). Designing synthetic tasks for human factors research: An application to uninhabited air vehicles. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 123-127). Santa Monica, CA: Human Factors and Ergonomics Society.

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science, 23,* 305-336.

Orasanu, J. (1990). Shared mental models and crew decision making. (Tech. Rep. No. 46). Princeton, NJ: Princeton University, Cognitive Science Laboratory.

Prince, C., Chidester, T. R., Cannon-Bowers, J., & Bowers, C. (1992). Aircrew coordination – Achieving teamwork in the cockpit. In R. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 329-353). New York: Ablex.

Prince, C. & Salas, E. (1993). Training and research for teamwork in the military aircrew. In E. L. Wiener, B. G. Kanki, and R. L. Helmreich (Eds.), *Cockpit resource management* (pp. 337-366). San Diego, CA: Academic Press.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology, 8,* 197-208.

Salas, E. Cannon-Bowers, J.A., Church-Payne, S., & Smith-Jentsch, K. A. (1998). Teams and teamwork in the military. In C. Cronin (Ed.), *Military Psychology: An Introduction* (pp. 71-87). Needham Heights, MA: Simon & Schuster.

Salas, E. Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3-29). Norwood, NJ: Ablex.

Stout, R., Cannon-Bowers, J. A., & Salas, E. (1996). The role of shared mental models in developing team situation awareness: Implications for training. *Training Research Journal, 2,* 85-116.

Stout, R. J., Salas, E., & Carson, R. (1994). Individual task proficiency and team process: What's important for team functioning? *Military Psychology, 6,* 177-192.

Schvaneveldt, R. W., Durso, F. T., Goldsmith, T. E., Breen, T. J., Cooke, N. M., Tucker, R. G., & DeMaio, J. C. (1985). Measuring the structure of expertise. *International Journal of Man-Machine Studies, 23,* 699-728.

Wickens, C. D. (1998). Commonsense statistics. *Ergonomics in Design, 6,* 18-22.

Zalesny, M. D., Salas, E., & Prince, C. (1995). Conceptual and measurement issues in coordination: Implications for team behavior and performance. In G. R. Ferris (Ed.), *Research in Personnel and Human Resources Management (Vol. 1, pp. 81-115).* Greenwich, CT: JAI Press.

PUBLICATIONS ASSOCIATED WITH THIS EFFORT

The following papers and presentations were supported or partially supported by this effort.

MEASURING TEAM COGNITION

Cooke, N. J., Stout, R., & Salas, E. (1997). Cognitive task analysis for team tasks. Paper presented at the ONR/NATO Workshop on Cognitive Task Analysis, October 30-November 1, 1997, Washington, D. C.

Cooke, N. J., Stout, R., Rivera, K., & Salas, E. (1998). Exploring measures of team knowledge. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*, 215-219 and paper presented at the 42nd annual meeting of the Human Factors and Ergonomics Society, October 5-9, Chicago, IL.

Cooke, N. J. (1999). Knowledge metrics for teams. Paper presented at Meeting of the Southwestern Psychological Association, April 1-3, Albuquerque, NM.

Cooke, N. J., Salas, E., Cannon-Bowers, J. A., & Stout, R. (2000). Measuring team knowledge. *Human Factors, 42*, 151-173.

CERTT (COGNITIVE ENGINEERING RESEARCH FOR TEAM TASKS) LABORATORY

Cooke, N. J. & Shope, S. M. (1998). Facility for Cognitive Engineering Research on Team Tasks. Technical Report for Grant No. F49620-97-1-0149, submitted to AFOSR, Bolling AFB, Washington, DC.

Cooke, N. J., & Shope, S. M. (1999). *CERTT Lab Video*. Produced by NMSU's Instructional Video Services. Funded by NMSU Department of Psychology, NMSU College of Arts and Sciences Research Center, and Sandia Research Corporation.

Cooke, N. J., & Rivera, K. (1999). *CERTT Lab Brochure*. Funded by NMSU Department of Psychology, NMSU College of Arts and Sciences Research Center, and Sandia Research Corporation.

Cooke, N. J. (1999). CERTT Lab. Poster presented at the technical group meeting of the Cognitive Engineering and Decision Making technical group at the 43rd annual meeting of the Human Factors and Ergonomics Society, September 27-October 1, Houston, TX.

UAV-STE (UNINHABITED AIR VEHICLES-SYNTHETIC TASK ENVIRONMENT)

Cooke, N. J., Rivera, K., Shope, S.M., & Caukwell, S. (1999). A synthetic task environment for team cognition research. *Proceedings of the Human Factors and Ergonomics Society 43rd*

Annual Meeting, 303-307 and paper presented at the 43rd annual meeting of the Human Factors and Ergonomics Society, September 27-October 1, Houston, TX.

Cooke, N. J. & Shope, S. M. (1999). CERTT-UAV Task. Invited talk and demonstration presented at the Scaled Worlds Symposium, June 24-27, Athens, GA.

Cooke, N. J., Shope, S.M., & Rivera, K. (2000). Control of an uninhabited air vehicle: A synthetic task environment for teams. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* and demonstration presented at the 44th annual meeting of the Human Factors and Ergonomics Society and International Ergonomics Association, July 30-August 4, San Diego, CA.

Cooke, N. J., & Shope, S. M. (under review). Designing a synthetic task environment. In Elliot, L. Book from Scaled Worlds Symposium.

EMPIRICAL STUDIES OF TEAM COGNITION

Cooke, N. J., Cannon-Bowers, J. A., Kiekel, P. A., Rivera, K., Stout, R., and Salas, E. (2000). Improving team's interpositional knowledge through cross training. *Proceedings of the Human Factors and Ergonomics Society 44th Annual Meeting* and Paper presented at 44th annual meeting of the Human Factors and Ergonomics Society and International Ergonomics Association, July 30-August 4, San Diego, CA.

Cooke, N. J., Kiekel, P. A., & Helm E. (under review). Measuring Team Knowledge During Skill Acquisition of a Complex Task. Submitted to *International Journal of Cognitive Ergonomics: Special Issue on Information Acquisition and Learning.*

Cooke, N. J., Kiekel, P.A., Salas, E., Stout, R., Bowers, C, & Cannon-Bowers, J.(in preparation). Measuring team knowledge: A window to the cognitive underpinnings of team performance differences. To be submitted to *Group Dynamics.*

Appendix A
TEAM PROCESS QUESTIONS
FOR STUDY 1, Missions 1-6, 8, 9

The following nine behaviors may or may not occur at the designated event triggers. Circle yes or no depending on whether or not the behavior occurred. (Event triggers are in italics.) Note: Both experimenters should independently complete these process ratings until some reliability has been established.

BEGINNING OF MISSION

P1 yes no

In first 5 minutes of mission team makes planning statements.

LVN-OAK OR FIRST ROZ BOX

P2 yes no

Prior to UAV in effective radius (within 5 miles of) of H-AREA or F-AREA or targets within first ROZ BOX DEMPC communicates restrictions on H-AREA and/or F-AREA to AVO.

P3 yes no

Prior to UAV in effective radius (within 5 miles of) of H-AREA or F-AREA or targets within first ROZ BOX AVO acknowledges the DEMPC's communication or requests the information.

P4 yes no

Prior to UAV in effective radius (within 5 miles of) of H-AREA or F-AREA or targets within first ROZ BOX DEMPC communicates upcoming targets (H-AREA, F-AREA) to PLO.

P5 yes no

Prior to UAV in effective radius (within 5 miles of) of H-AREA or F-AREA or targets within first ROZ BOX PLO acknowledges the DEMPC's communication or requests the information.

AFTER KGM-FRT CALL-IN

P6 yes no

Within 5 minutes after call-in of new ROZ box (KGM-FRT) DEMPC communicates new ROZ (KGM-FRT) and new targets to AVO and PLO

PRK-ASH OR SECOND ROZ BOX

P7 yes no

Prior to UAV in effective radius (within 5 miles of) of S-STE or MSTE or targets within second ROZ box DEMPC anticipates PLO's need and communicates the PRK-ASH targets (S-STE, MSTE) without PLO asking.

P8 yes no

While UAV within PRK-ASH ROZ box (e.g., 2.5 miles of PRK or ASH) or within second ROZ box AVO and PLO work together to maneuver UAV for photos (this should be evident in their communication).

END OF MISSION

P9 yes no

Within 5 minutes after end of mission team assesses and discusses their performance.

Other:

Please note any other behaviors that were indicative of good or poor team process behaviors.

Appendix A
TEAM PROCESS QUESTIONS
FOR STUDY 1, Missions 7 & 10

The following nine behaviors may or may not occur at the designated event triggers. Circle yes or no depending on whether or not the behavior occurred. (Event triggers are in italics.) Note: Both experimenters should independently complete these process ratings until some reliability has been established.

BEGINNING OF MISSION

P1 yes no

In first 5 minutes of mission team makes planning statements.

MAR-MON OR FIRST ROZ BOX

P2 yes no

Prior to UAV in effective radius (within 5 miles of) of SAN or TKE or targets within first ROZ BOX DEMPC communicates restrictions on SAN and/or TKE to AVO.

P3 yes no

Prior to UAV in effective radius (within 5 miles of) of SAN or TKE or targets within first ROZ BOX AVO acknowledges the DEMPC's communication or requests the information.

P4 yes no

Prior to UAV in effective radius (within 5 miles of) of SAN or TKE or targets within first ROZ BOX DEMPC communicates upcoming targets (SAN, TKE) to PLO.

P5 yes no

Prior to UAV in effective radius (within 5 miles of) of SAN or TKE or targets within first ROZ BOX PLO acknowledges the DEMPC's communication or requests the information.

BYU-DC10 OR SECOND ROZ BOX

P6 yes no

Prior to UAV in effective radius (within 5 miles of) of SP or WIC or targets within second ROZ box DEMPC anticipates PLO's need and communicates the BYU-DC10 targets (SP, WIC) without PLO asking

P7 yes no

While UAV within BYU-DC10 ROZ box (e.g., 2.5 miles of BYU or DC10) or within second ROZ box AVO and PLO work together to maneuver UAV for photos (this should be evident in their communication).

AFTER WP30-BAY CALL-IN

P8 yes no

Within 5 minutes after call-in of new ROZ box (WP30-BAY) DEMPC communicates new ROZ (WP30-BAY) and new targets to AVO and PLO

END OF MISSION

P9 yes no

Within 5 minutes after end of mission team assesses and discusses their performance.

Other: Please note any other behaviors that were indicative of good or poor team process behaviors.

Appendix A
TEAM PROCESS QUESTIONS FOR STUDY 2

The non-talking experimenter will evaluate the team process. The following behaviors may or may not occur at the designated event triggers. Circle yes or no depending on whether or not the behavior occurred.

BEGINNING OF MISSION

P1 **yes** **no**

Within one minute prior to the start of the mission and the team reaching the effective radius of the first target, the TEAM discusses how they will perform during the mission.

LVN-OAK ROZ BOX

P2 **yes** **no**

Prior to UAV in effective radius of H-AREA or F-AREA, AVO acknowledges the DEMPC's communication or requests the information.

AFTER KGM-FRT CALL-IN

P3 **yes** **no**

Within 5 minutes after call-in of new ROZ box, DEMPC communicates new ROZ and new targets to AVO and PLO.

PRK-ASH ROZ BOX

P4? **yes** **no**

Prior to UAV in effective radius of S-STE or MSTE, PLO asks for PRK-ASH targets before being told by the DEMPC.

P5 **yes** **no**

While UAV within PRK-ASH ROZ box (e.g., 2.5 miles of PRK or ASH), AVO and PLO work together to maneuver UAV for photos (this should be evident in their communcation).

END OF MISSION

P6 **yes** **no**

Within 5 minutes after end of mission, the TEAM assesses and discusses their performance.

Other:

Please note any other behaviors that were indicative of good or poor team process behaviors at the experimenter station.

Appendix B
SITUATION AWARENESS QUERIES FOR STUDY 1

The SA questions will be called in to individuals by you, the experimenter who plays the role of intelligence requesting information. Only one experimenter should communicate with the team for the two call-ins and the three SA queries. Each of the following three questions is posed to each of the three team members individually and responses are made individually by each member in turn. The queries will each be delivered to the three team members in a predetermined random order during a five-minute interval. The five-minute interval will also be determined randomly for each mission. Move quickly during the five minutes so that all team members can respond to approximately the same question. Be sure to record the team members responses along with the actual situation outcome.

Because we want to avoid team members broadcasting their responses to these queries to other team members, you should preface each query (and the 2 call-ins) with the following statement:

Intelligence calling _____. I have a request for information. Be sure your intercom switches are all down except for EXP so that your response can be kept top secret.

RANDOMIZATION

1) Randomly select three of the following 7 intervals. Do not repeat an interval. Mark the order of the three selected intervals below.

Number	**Minutes**	**Selection Order (1, 2, or 3)**
1	5-10	___
2	10-15	___
3	15-20	___
4	20-25	___
5	25-30	___
6	30-35	___
7	35-40	___

2) Also before each mission use a random numbers table (in appendix of manual) to randomly select three of the following 6 team member orders. Do not repeat an order. Mark the order of the three selected orders below.

Number	**Minutes**	**Selection Order (1, 2, or 3)**
1	APD	___
2	ADP	___
3	PAD	___
4	PDA	___
5	DAP	___
6	DPS	___

*A=AVO, P=PLO, D=DEMPC

3) Copy the minutes and orders into the appropriate spaces below (e.g., the minutes and order labelled 1 is placed in the blanks under SA-A, ominutes and order 2 under SA-B etc.) Minutes indicates the interval during which this question will be asked. Order indicates the team member who will be asked first, second, and third.

SA-A
Minutes ___
Order ______
Intelligence calling _____. I have a request for information. Be sure your intercom switches are all down except for EXP so that your response can be kept top secret. We are trying to prepare another UAV team for the next mission. How many targets do you think your team manage to successfully photograph by the end of your 40-minute mission? There are nine targets total.

	Response (0-9)
AVO	___
PLO	___
DEMPC	___
Actual	___

SA-B
Minutes ___
Order ______
Intelligence calling _____. I have a request for information. Be sure your intercom switches are all down except for EXP so that your response can be kept top secret. We have detected a minor communication system fault. In order to help repair the problem can you tell me with which team member or members you plan to communicate next and the topic of that communication?

RESPONSES

	WILL TALK WITH (circle one or both)	**TOPIC**
AVO	PLO DEMPC	____________
PLO	AVO DEMPC	____________
DEMPC	AVO PLO	____________

ACTUAL

	TALKED WITH (circle one or both)	**TOPIC**
AVO	PLO DEMPC	____________
PLO	AVO DEMPC	____________
DEMPC	AVO PLO	____________

SA-C
Minutes ___
Order ______
Intelligence calling _____. I have a request for information. Be sure your intercom switches are all down except for EXP so that your response can be kept top secret. For intelligence purposes can you tell me how many targets has your team successfully photographed in this mission thus far? There are nine targets total.

	Response (0-9)
AVO	___
PLO	___
DEMPC	___
Actual	___

Exp. AF00-02
Team ID _______
Date___________
Mission 1 2 3 4 5
~~Experimenter~~ ______

Appendix B
SITUATION AWARENESS QUERIES FOR STUDY 2

The SA questions will be called in to the team only by the talking experimenter playing the role of Intelligence. Each of the following three questions is posed to each of the three team members individually and responses are made individually by each member in turn. Discourage them from asking one another the answers to the questions. The queries will each be delivered to the three team members in a predetermined random order during a five-minute interval. The five-minute interval will also be determined randomly for each mission. Move quickly during the five minutes so that all team members can respond to approximately the same question. Be sure to record the team members responses along with the actual situation outcome.

RANDOMIZATION

1) Randomly select three of the following 7 intervals by drawing numbers out of a hat. Do not repeat an interval. Mark the order of the three selected intervals below.

Number	Minutes	Selection Order (1, 2, or 3)
1	5-10	___
2	10-15	___
3	15-20	___
4	20-25	___
5	25-30	___
6	30-35	___
7	35-40	___

2) Randomly select three of the following 6 team member orders by drawing numbers out of a hat. Do not repeat an order. Mark the order of the three selected orders below.

Number	Minutes	Selection Order (1, 2, or 3)
1	APD	___
2	ADP	___
3	PAD	___
4	PDA	___
5	DAP	___
6	DPA	___

*A=AVO, P=PLO, D=DEMPC

3) Copy the minutes and orders into the appropriate spaces below (e.g., the minutes and order selected first is placed in the blanks under SA-A, minutes and order selected second under SA-B, etc.) Minutes indicates the interval during which this question will be asked. Order indicates the team member who will be asked first, second, and third.

SA-A
Minutes ______ **Order**________

Mission 1: This is intelligence calling the ____. I have a request for information. Please be sure that all of your communication switches are down except for EXP so that your responses can be kept top secret. We are trying to

prepare another UAV team for the next mission. How many targets do you think your team will manage to successfully photograph by the end of your 40-minute mission? There are nine targets total.
Mission 2 – 5: This is intelligence calling the ____. *(Experimenter: watch to be sure only EXP switch is up when he/she responds.)* How many targets do you think your team will manage to successfully photograph by the end of your 40-minute mission?

Response: AVO_____ PLO_____ DEMPC _____ Actual______

SA-B
Minutes ______ Order_______

Mission 1: This is intelligence calling the ____. I have a request for information. Please be sure that all of your communication switches are down except for EXP so that your responses can be kept top secret. For intelligence purposes can you tell me how many targets your team has successfully photographed in this mission thus far? There are nine targets total.
Mission 2 – 5: This is intelligence calling the ____. *(Experimenter: watch to be sure only EXP switch is up when he/she responds.)* Can you tell me how many targets your team has successfully photographed in this mission thus far?

Response: AVO_____ PLO_____ DEMPC _____ Actual______

SA-C
Minutes ______ Order_______

Mission 1: This is intelligence calling the ____. I have a request for information. Please be sure that all of your communication switches are down except for EXP so that your responses can be kept top secret. For intelligence purposes can you tell me which target you are scheduled to photograph next? *(Experimenter: if the TO WAYPOINT is a target, ask them which target is next after this one?)*
Mission 2 – 5: This is intelligence calling the ____. *(Watch to be sure only EXP switch is up when he/she responds.)* Can you tell me which target you are scheduled to photograph next? *(Experimenter: if the TO WAYPOINT is a target, ask them which target is next after this one?)*

Response: AVO_____ PLO_____ DEMPC _____ Actual______

Experimenter: To the following questions, answer **1** for **YES** and **0** for **NO** in regards to SA-C.

AVO's response same as PLO's?		AVO's response same as Actual?	
AVO's response same as DEMPC's?		PLO's response same as Actual?	
PLO's response same as DEMPC's?		DEMPC's response same as Actual?	

Appendix C
TASKWORK QUESTIONS FOR STUDY 1

Please answer the following questions as accurately and honestly as you can. Complete sentences are not necessary and brevity is appreciated.

PART 1:
What is the main goal or objective of the UAV task?

____

PART 2:
Now think of how this goal is accomplished. What are the subgoals (in any order) that are required to achieve this goal? List at least 2, but no more than 4.

1. ____________________
2. ____________________
3. ____________________
4. ____________________

PART 3
Next, write each subgoal listed in Part 2 above on the appropriate line of the worksheet. Then in the first column under each subgoal list the tasks (in any order) required to complete this subgoal. List at least 2, but no more than 6 per subgoal. Ignore the columns marked "A" and "B" for now.

PART 4:
Next, consider who performs each of the tasks that you just identified. In the worksheet column labeled "A" list the team member(s) responsible for this task. The roles are AVO, PLO, or DEMPC.

PART 5:
For each subgoal, consider the sequence, if any, in which the tasks have to be carried out. Does one task have to occur, before another can be done? Please indicate sequence information by numbering each task in the worksheet column labeled "B". Use number "1" for the first task in a sequence, "2" for the second, etc. If several tasks occur simultaneously number them the same. If sequence is not an issue for a task, leave that row blank.

WORKSHEET

SUBGOAL 1: ____________	(A)	(B)
Task 1: ________________	________	________
Task 2: ________________	________	________
Task 3: ________________	________	________
Task 4: ________________	________	________
Task 5: ________________	________	________
Task 6: ________________	________	________
SUBGOAL 2: ____________	(A)	(B)
Task 1: ________________	________	________

Task 2: ________________	________	________
Task 3: ________________	________	________
Task 4: ________________	________	________
Task 5: ________________	________	________
Task 6: ________________	________	________
<u>SUBGOAL 3:</u> ____________	(A)	(B)
Task 1: ________________	________	________
Task 2: ________________	________	________
Task 3: ________________	________	________
Task 4: ________________	________	________
Task 5: ________________	________	________
Task 6: ________________	________	________
<u>SUBGOAL 4:</u> ____________	(A)	(B)
Task 1: ________________	________	________
Task 2: ________________	________	________
Task 3: ________________	________	________
Task 4: ________________	________	________
Task 5: ________________	________	________
Task 6: ________________	________	________

SCORING INFORMATION FOR TASKWORK QUESTIONNAIRE

Part 1
There was only one correct answer – to take photographs. This was scored as 1 or 0 and was a proportion of the number of goals listed if there were more than one. Agreement among pairs of team members was also computed.

Part 2
If it helped the team to accomplish photo-taking t is scored as correct (e.g., "enter target area"). The score was the proportion of correct items over the total number of items listed. Agreement among team members was computed as the number of items agreed upon overall divided by the number of individual items listed for that particular pair of the three team members.

Part 3
Correctness was determined based on the subgoal and the scorer's knowledge of the task. Again correctness was scored and proportion correct responses out of the total number of subgoals listed was recorded. Agreement among pairs of team members was also computed, but was a rare occurrence.

Part 4
Only correctness was scored and this was based on the scorer's knowledge of the task. This was scored as the total number of correct team position assignments divided by the total number of team position assignments.

Part 5
Not scored

Team ID ______
ROLE: A P D
Date ______
Know. Sess. 1 2 3
Experimenter ___

Appendix C
TEAMWORK QUESTIONS
FOR STUDY 1

<u>**Scenario:**</u> **Imagine that you are in the middle of a UAV mission. During that mission, a target waypoint is coming up that is passable, but that is associated with some kind of cloud cover. Now consider the communications that occur between your team members while you answer the following questions.**

<u>**PART 1:**</u>

1) In the context of this scenario to whom does the AVO pass information? Check all that apply. If there is no information passed by the AVO to that individual leave that space blank.

The AVO informs...

___the PLO
___the DEMPC

2) In the context of this scenario to whom does the PLO pass information? Check all that apply. If there is no information passed by the PLO to that individual leave that space blank.

The PLO informs ...

___the AVO
___the DEMPC

3) In the context of this scenario to whom does the DEMPC pass information? Check all that apply. If there is no information passed by the DEMPC to that individual leave that space blank.

The DEMPC informs ...

___the AVO
___the PLO

PART 2:

Now for each time in Part 1 that you checked that one teammate informs another, describe the nature of that information. Remember to answer this in the context of the scenario described above. If you did not check the particular alternative leave it blank.

___The AVO informs the PLO about ______________________________.

___The AVO informs the DEMPC about __________________________.

___The PLO informs the AVO about ______________________________.

___The PLO informs the DEMPC about__________________________.

___The DEMPC informs the AVO about ___________________________.

___The DEMPC informs the PLO about ___________________________.

PART 3:

Now consider the timing of information. Still in the context of the scenario, does some information have to come before other information? For each time information is passed, identify its sequential order by placing a number in the leftmost column indicating order. A 1 means that information comes first in the sequence, a 2 means that it comes second, etc. If several pieces of information are passed simultaneously give them the same number. If sequence is not an issue for a specific piece of information then leave that row blank.

SCORING INFORMATION FOR THE TEAMWORK QUESTIONNAIRE

Part 1
Accuracy was not scored. Agreement was scored as an average for each pair of team members of the number of selections agreed upon out of a total of 6.

Part 2
Accuracy was scored by the scorers' knowledge of the task and determining whether a response was correct or incorrect. Score was based on number correct out of the total number of items listed. Agreement was scored as answers agreed upon over the total number of answers.

Part 3
This was scored for accuracy based on the scorer's understanding of the task. Agreement was the number of sequences given that matched over the total number of sequences given.

A composite team accuracy score was created by summing the proportion of accuracy for each team member on Parts 2 and 3. Part 2 accuracies were multiplied by two in order to weigh that type of accuracy more heavily. The entire sum of proportions was divided by 9 (a perfect team accuracy score).

Appendix C
TEAMWORK QUESTIONS FOR STUDY 2

Scenario: Imagine that you are in the middle of a UAV mission. During that mission, a non-priority target waypoint is coming up that is passable, but due to cloud cover, there are restrictions on the airspeed and altitude. Now consider the communications that will occur between you and your team members while you answer the following questions.

Part 1: In the context of this scenario, decide which team members you think will pass information and who will receive information. In the Part 1 column, check the boxes to the left of who you believe will send and receive information. Example: If in this scenario you think the AVO will pass information to the PLO, check the box next to "The AVO informs the PLO." If you **don't** think the AVO will pass information to the PLO, leave that box empty. Check as many or as few boxes as you like.

Part 2: For the boxes that you checked in Part 1, describe on the lines in the Part 2 column the nature of the information the sender passes to the receiver. For the boxes that you left empty, also leave the lines in Part 2 empty.

Part 3: Now decide in what order the information in Part 2 is passed. In the blanks in the Part 3 column, number from one to however many boxes you checked the order in which the information is passed. A "1" means that piece of information must be passed first, a "2" means that piece of information is passed second, an so on. If several pieces of information are passed simultaneously give them the same number. If sequence is not an issue for a specific piece of information then leave that space blank.

Scenario: Imagine that you are in the middle of a UAV mission. During that mission, a non-priority target waypoint is coming up that is passable, but due to cloud cover, there are restrictions on the airspeed and altitude. Now consider the communications that will occur between you and your team members while you answer the following questions.

Part 1: Check the boxes for who you think will pass information to whom.
Part 2: Write down what information will be passed for each box you checked.
Part 3: Number the sequence in which that information will be passed.

Part 1	Part 2	Part 3
☐ The AVO informs the PLO about	______________________	______
☐ The AVO informs the DEMPC about	______________________	______
☐ The PLO informs the AVO about	______________________	______
☐ The PLO informs the DEMPC about	______________________	______
☐ The DEMPC informs the AVO about	______________________	______
☐ The DEMPC informs the PLO about	______________________	______

SCORING INFORMATION FOR THE TEAMWORK QUESTIONNAIRE

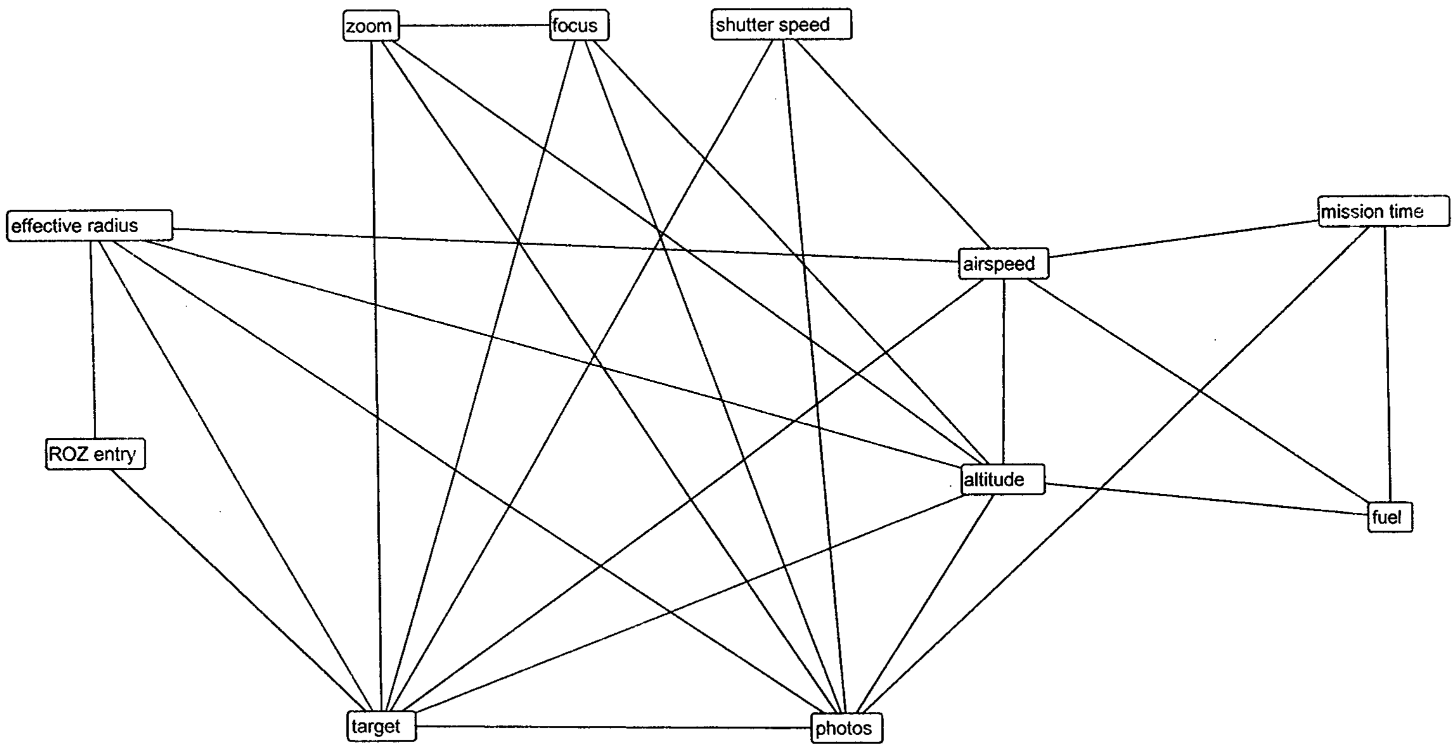
Part 1
Accuracy was not scored. Agreement was scored as an average for each pair of team members of the number of selections agreed upon out of a total of 6.

Part 2
Accuracy was scored by the scorers' knowledge of the task and determining whether a response was correct or incorrect. Score was based on number correct out of the total number of items listed. Agreement was scored as correct answers agreed upon over the total number of answers given and was given more weight in the composite agreement score.

Part 3
Not scored for accuracy. Agreement was the number of sequences given that matched over the total number of sequences given.

Appendix D
REFERENT PATHFINDER NETWORKS USED TO SCORE TASKWORK ACCURACY

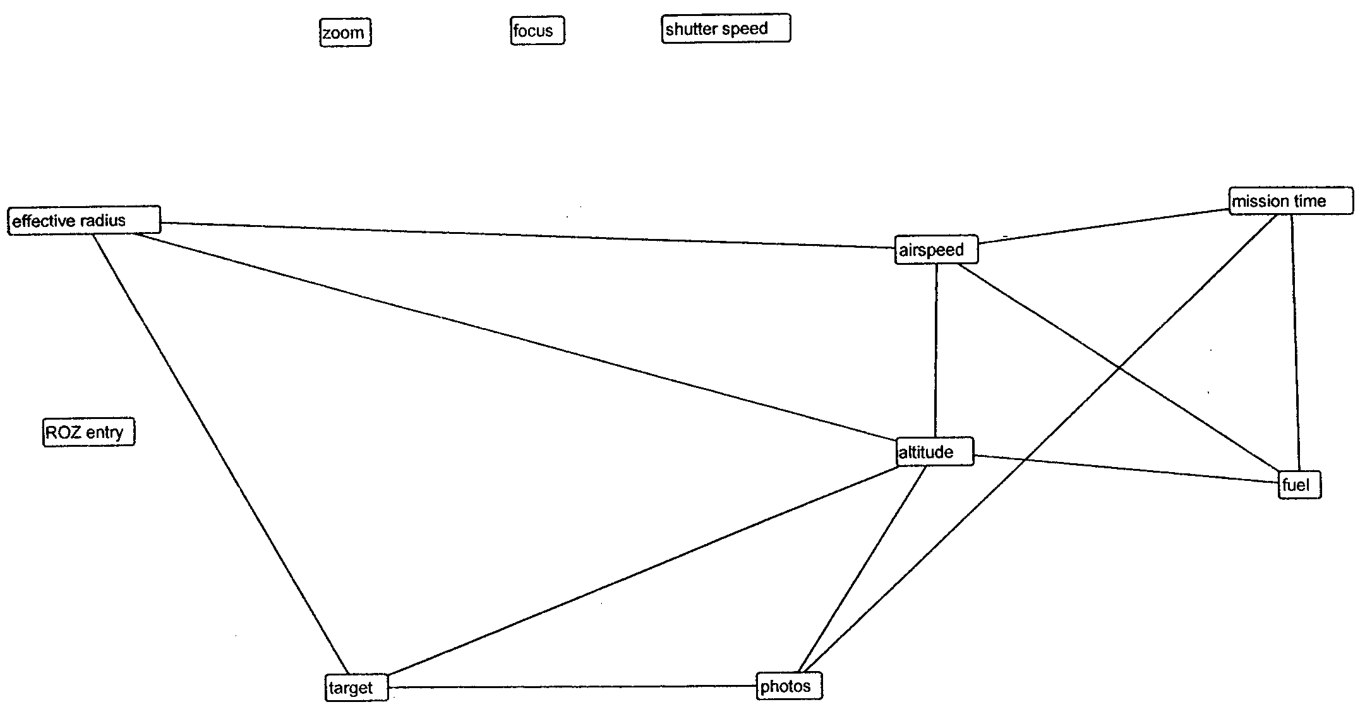
OVERALL REFERENT



Appendix D

REFERENT PATHFINDER NETWORKS USED TO SCORE TASKWORK ACCURACY

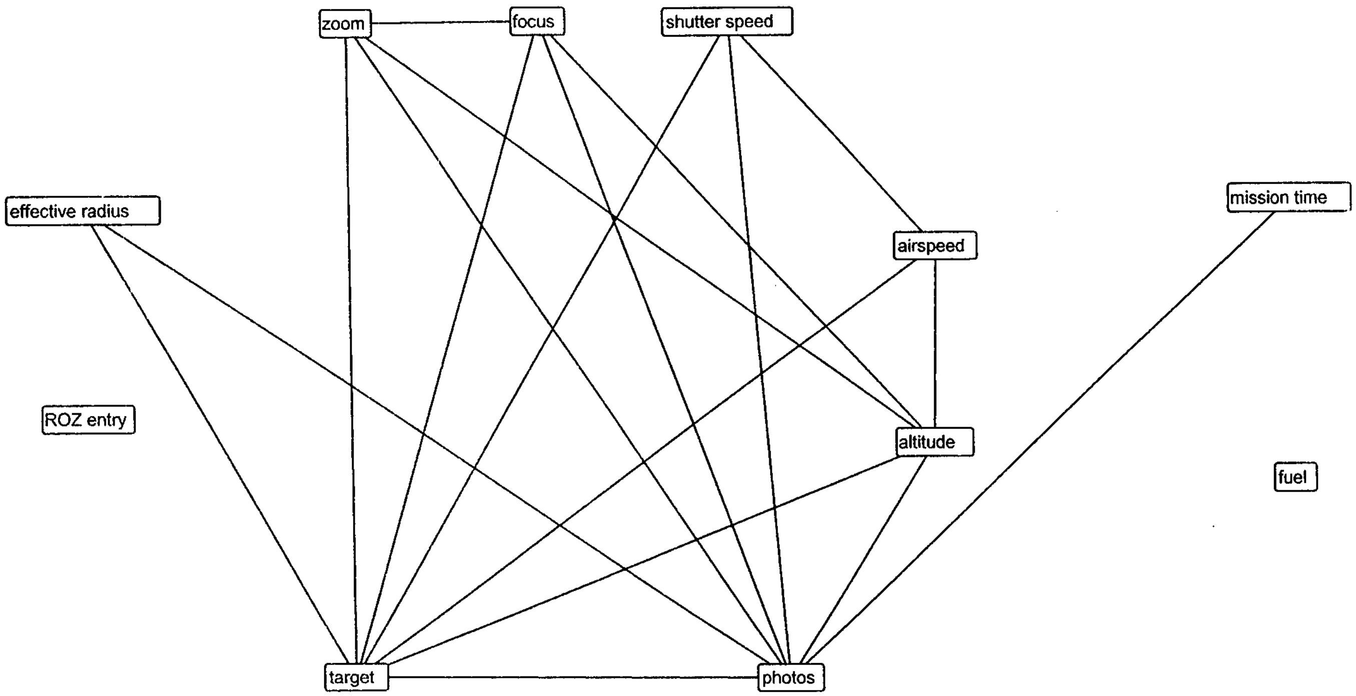
AIR VEHICLE OPERATOR REFERENT



Appendix D

REFERENT PATHFINDER NETWORKS USED TO SCORE TASKWORK ACCURACY

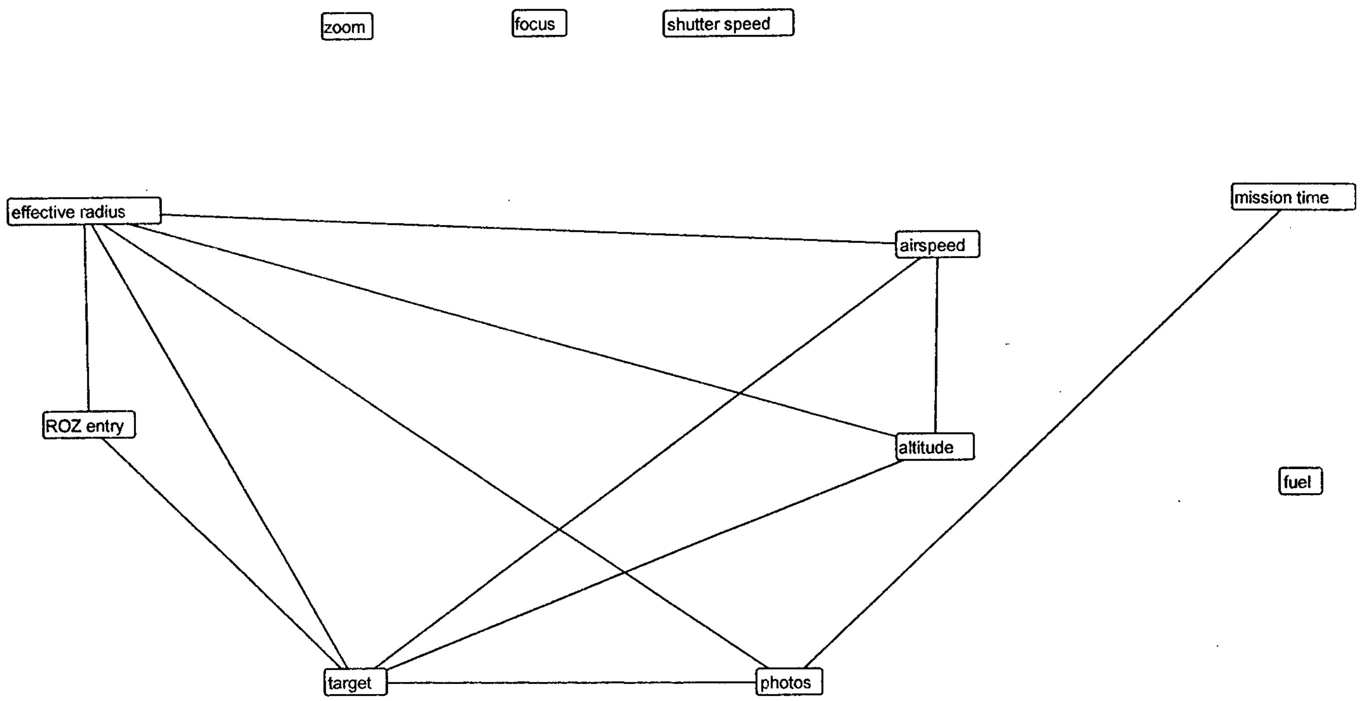
PAYLOAD OPERATOR REFERENT



Appendix D

REFERENT PATHFINDER NETWORKS USED TO SCORE TASKWORK ACCURACY

DEMPC REFERENT



Appendix E
BASIC SKILLS CHECKLIST

Have the following behaviors performed by the three team members in order and check them off as they are accomplished. With two experimenters, the DEMPC and AVO checks can be conducted in parallel with the PLO checks following.

COMMUNICATION CHECKS

Everyone should put headsets on, including the experimenters. Experimenters talk to team members over the headsets conducting the following checks. Adjust microphones and instruct on push-to-talk button and intercom as needed.

Put all intercom switches down except for experimenter. Experimenter queries each team member in turn:

___1. Experimenter can hear AVO
___2. AVO can hear Experimenter
___3. Experimenter can hear PLO
___4. PLO can hear experimenter
___5. Experimenter can hear DEMPC
___6. DEMPC can hear experimenter

All intercom switches up. Experimenter queries each team member in turn:
___7. Experimenter can hear everyone
___8. AVO can hear PLO and DEMPC
___9. PLO can hear AVO and DEMPC
___109. DEMPC can hear AVO and PLO

All intercom switches down. Instruct team members to flip appropriate switch up to talk.
___11. AVO can talk to DEMPC only
___12. PLO can talk to AVO only
___13. DEMPC can talk to PLO only

Remove and stow headsets. Start the UAV simulation. Ask the team members to do each of the following activities and check them off as they are observed.

DEMPC CHECKS

___14. DEMPC can hide map detail and waypoints and can show detail and waypoints.

___15. Delete waypoint WLF from the flight plan

___16. Insert waypoint HJ10 into the flight plan between ROC and F-AREA

___17. Sequence the plan until the following subset of 5 is highlighted: CMN, H-AREA, ROC, HJ10, F-AREA.
___18. Send this route

AVO CHECKS

___19. Adjust course so that you are heading to the "To Waypoint," DEX. Keep adjusting course throughout checks to minimize deviation.

___20. Increase airspeed to slightly above 200 (do not need to wait until speed, etc. gets to setting).

___21. Increase altitude to 3200

___22. Raise flaps and landing gear

___23. Increase speed above acceptable range and observe the warning and alarm.

___24. Change the queued waypoint to H-AREA.

___25. Make H-AREA the new "To Waypoint"

___26. Adjust course to head toward H-AREA. Keep adjusting course throughout checks to minimize deviation.

___27. Decrease airspeed to around 80 knots.

___28. Refuel

Keep adjusting course to head toward H-AREA maintaining current airspeed and altitude. This is necessary for the PLO checks.

PLO CHECKS

___29. The upcoming waypoint H-AREA is a target (a hangar). The effective radius is 5 miles. Find the photo requirements for this target.

___30. Set the camera settings.

___31. Take a picture. If it is good press accept. If it is not keep adjusting settings until it is.

___32. Change a setting and take a photo that is not good.

___33. Reload the film.

Appendix F
DEBRIEFING QUESTIONS FOR STUDY 1

First Screen:

Enter Participant ID

First Page:

1. Gender?

    Male
    Female

2. Class?

    Freshman
    Sophomore
    Junior
    Senior
    Grad Student

3. Enter current GPA:

4. Enter major:

Second Page:

5. I enjoyed participating in this study.

6. I enjoyed the team task part of this study.

7. I would welcome the opportunity to participate in a similar study in the future.

8. My team worked well together.

Third Page:

9. I performed well on this task.

10. I would like to work with my fellow team members again.

11. I like playing video and computer games..

12. I like to be part of a team.

Fourth Page:

14. The person that played the AVO role was:

    Me
    A stranger to me
    Somewhat familiar
    Well known to me

15. The person that played the PLO role was:

    Me
    A stranger to me
    Somewhat familiar
    Well known to me

16. The person that played the DEMPC role was:

    Me
    A stranger to me
    Somewhat familiar
    Well known to me

Submit Responses.

Appendix F
DEBRIEFING QUESTIONS FOR STUDY 2

First Screen:

What is your job? (AVO, PLO, DEMPC)

First Page:

1. Select rank:

   Cadet 4^{th} Class
   Cadet 3^{rd} Class
   Cadet Private
   Cadet Private 1^{st} Class
   Cadet Corporal
   Cadet Sergeant
   Cadet Staff Sergeant
   Cadet Sergeant 1^{st} Class
   Cadet Master Sergeant
   Cadet First Sergeant
   Cadet Staff Sergeant Major
   Cadet Sergeant Major
   Cadet 2^{nd} Lieutenant
   Cadet 1^{st} Lieutenant
   Cadet Captain
   Cadet Major
   Cadet Lt. Colonel
   Cadet Colonel
   Cadet Airman
   Cadet Airman 1^{st} Class
   Cadet Senior Airman
   Cadet Technical Sergeant
   Cadet Senior Master Sergeant
   Cadet First Sergeant
   Cadet Chief Master Sergeant
   Cadet Command Chief Master Sergeant

2. Select Major:

   All others not listed
   Accounting
   Agricultural Biology
   Agricultural Ecology….

3. Highest Aviation Training:

No aviation training
Ground school
Flight training
Private pilot certificate
Advanced pilot certificate

4. Select Ethnicity

African American
Asian American
Caucasian
Hispanic
Native American
Other

5. Class

Freshman
Sophomore
Junior
Senior
Graduate Student

6. Gender

Male
Female

7. Enter Current GPA:

Second Page:

8. I enjoyed participating in this study.
9. I enjoyed the team task part of this study.
10. I would welcome the opportunity to participate in this study in the future.
11. I would like to work with my fellow team members again.
12. I like playing video and computer games.
13. I like to be part of a team.
14. I was a successful member of the team.
15. My team worked well together.
16. I performed well on this task.
17. The AVO was competent.
18. The AVO contributed to the team.
19. The AVO tried hard.
20. The AVO was lucky.

21. The AVO had an easy task.
22. The AVO was likeable.
23. The PLO was competent.
24. The PLO contributed to the team.
25. The PLO tried hard.
26. The PLO was lucky.
27. The PLO had an easy task.
28. The PLO was likeable.
29. The DEMPC was competent.
30. The DEMPC contributed to the team.
31. The DEMPC tried hard.
32. The DEMPC was lucky.
33. The DEMPC had an easy task.
34. The DEMPC was likeable.
35. My team performed well on this task.

Third Page:

36. The person that played the AVO role was:

    Me
    A stranger to me
    Somewhat familiar
    Well known to me

37. The person that played the PLO role was:

    Me
    A stranger to me
    Somewhat familiar
    Well known to me

38. The person that played the DEMPC role was:

    Me
    A stranger to me
    Somewhat familiar
    Well known to me

Fourth Page:

39. Please answer in the box below:

Describe the strategies that you and your fellow team members used to generate conceptual relatedness ratings at the team level, given discrepancies among two or three individual ratings for that concept pair.

Thank you, please press ok and wait for instructions from the experimenter.